

Article

A Neighborhood Rough Sets-Based Attribute Reduction Method Using Lebesgue and Entropy Measures

Lin Sun ¹, Lanying Wang ¹, Jiucheng Xu ^{1,*} and Shiguang Zhang ^{1,2}

¹ College of Computer and Information Engineering, Henan Normal University, Xinxiang 453007, Henan, China; linsunok@gmail.com (L.S.); wanglanying5@126.com (L.W.); sg201106@163.com (S.Z.)

² School of Computer Science and Technology, Tianjin University, Tianjin 300072, China

* Correspondence: jiuchxu@gmail.com

Received: 15 December 2018; Accepted: 30 January 2019; Published: 1 February 2019



Abstract: For continuous numerical data sets, neighborhood rough sets-based attribute reduction is an important step for improving classification performance. However, most of the traditional reduction algorithms can only handle finite sets, and yield low accuracy and high cardinality. In this paper, a novel attribute reduction method using Lebesgue and entropy measures in neighborhood rough sets is proposed, which has the ability of dealing with continuous numerical data whilst maintaining the original classification information. First, Fisher score method is employed to eliminate irrelevant attributes to significantly reduce computation complexity for high-dimensional data sets. Then, Lebesgue measure is introduced into neighborhood rough sets to investigate uncertainty measure. In order to analyze the uncertainty and noisy of neighborhood decision systems well, based on Lebesgue and entropy measures, some neighborhood entropy-based uncertainty measures are presented, and by combining algebra view with information view in neighborhood rough sets, a neighborhood roughness joint entropy is developed in neighborhood decision systems. Moreover, some of their properties are derived and the relationships are established, which help to understand the essence of knowledge and the uncertainty of neighborhood decision systems. Finally, a heuristic attribute reduction algorithm is designed to improve the classification performance of large-scale complex data. The experimental results under an instance and several public data sets show that the proposed method is very effective for selecting the most relevant attributes with high classification accuracy.

Keywords: rough sets; neighborhood rough sets; attribute reduction; neighborhood entropy; Lebesgue measure

1. Introduction

Over the past few decades, data classification has become one of the important aspects of data mining, machine learning, pattern recognition, etc. As an important application of rough set models in a variety of practical problems, attribute reduction methods in information systems have been drawing wide attention of researchers [1,2]. It is a fundamental research theme in the field of granular computing [3]. Since a lot of information is gathered and it may include a large number of redundant and noisy attributes, the main objective of attribute reduction based on rough sets is to eliminate the redundant attributes, classify data and extract useful information [4].

Attribute reduction in rough set theory has been recognized as an important feature selection method [2]. Considering whether the evaluation criterion involves classification models, the existing feature selection methods can be broadly classified into the following three categories [5]: filter,

wrapper and embedded methods. Based on the intrinsic properties of data set, the filter methods select a feature subset as a preprocessing step that is independent of the learning algorithm [6]. Lyu et al. [7] investigated a filter method based on maximal information coefficient, which eliminates redundant information that does not require additional processes. The wrapper methods use a classifier to find the most discriminant feature subset by minimizing an error prediction function [8]. Jadhav et al. [9] designed a wrapper feature selection method and performed functional ranking based on information gain directed genetic algorithm. Unfortunately, the wrapper methods not only exhibit sensitivity to the classifier, but also tend to consume a lot of runtime [10]. Hence, few works in the field employ these methods. The embedded methods integrate feature selection in the training process to reduce the total time required for reclassifying different subsets [11]. Imani et al. [12] introduced an embedded algorithm based on a Chi-square feature selector. But it is not as accurate as the wrapper in classification problems. On the basis of the above analysis, our attribute reduction method is based on the filter method, in which a heuristic search algorithm is used to find an optimal attribute reduction subset for complex data sets using neighborhood rough set model.

Rough set theory has become an efficient mathematical tool for attribute reduction to discover data dependencies and reduce redundancy attributes contained in data sets [13,14]. Gu et al. [15] proposed a kernelized fuzzy rough set, but the result is critically depending on setting control parameters and the design of objective function. Raza and Qamar [16] presented a parallel rough sets-based dependency calculation method for feature selection, and it could directly find the positive region-based objects without calculating the positive region itself. However, rough sets can only deal with attributes of a specific type in information systems using a binary relation [17]. In addition, the traditional rough set model need to discretize the data when dealing with continuous data, but the process ignores the differences among data and affects the information expression of the original attribute set to a certain degree [18]. Moreover, the original property of the continuous-valued data will change after discretization, and some useful information will be lost [8]. To overcome this drawback, scholars have developed many extensions for the traditional rough set model [19,20]. As an extended rough set model, neighborhood rough set model is introduced to solve the problem that classical rough sets cannot handle continuous numerical data. Since most of data in attribute reduction are numerical, when utilizing neighborhood rough sets, the discretization of continuous data can be avoided [21]. Hu et al. [22] developed a neighborhood rough set model via the δ -neighborhood set to deal with discrete and continuous data sets. Chen et al. [13] proposed a neighborhood rough sets-based feature reduction fish swarm algorithm to deal with numerical data sets. Sun et al. [23] studied a gene selection algorithm based on Fisher linear discriminant and neighborhood rough sets, which is of great practical significance for cancer clinical diagnosis. Mu et al. [24] investigated a gene selection method using Fisher transformation based on neighborhood rough sets for numerical data sets. Li et al. [14] developed a feature reduction method based on neighborhood rough sets and discernibility matrix, but there is a hypothesis that all features data are available. Nonetheless, since the global neighborhood in this field is only used to deal with decision systems, that is, each sample uses the same neighborhood value in different conditional attribute combinations; this method has a high time complexity and does not result in the optimal δ value [24]. Wang et al. [1] constructed local neighborhood rough sets to deal with labeled data. It is well known that the neighborhood rough sets can be employed to deal with an information system with heterogeneous attributes including categorical and numerical attributes [22]. However, a large number of existing attribute reduction algorithms based on rough set model and its variations only analyze finite sets, which would limit their application to some extension. Halmos [25] used Lebesgue measure as measure theory to achieve uncertainty measures. Song and Li [26] stated Lebesgue theorem in non-additive measure theory. Xu et al. [27] introduced Lebesgue integral over infinite interval and presented a computation method based on kernel function for uncertainty measures. Halcinová et al. [28] investigated the weighted Lebesgue integral by Lebesgue differentiation theorem, and used Lebesgue measure to develop the standard weighted L_p -based sizes. Park et al. [29] expressed a measurable map through Lebesgue integration to define

the Cumulative Distribution Transform of density functions. Recently, scholars [30–32] introduced Lebesgue measure as a promising additional method to resolve some problems in the application of data analysis. As inspired by the ideas of Lebesgue measure that can efficiently evaluate infinite sets, it is therefore necessary to employ Lebesgue measure to study uncertainty measures and efficient reduction algorithms for infinite sets in neighborhood decision systems.

Uncertainty measures play an important role in the uncertainty analysis of granular computing models [33]. Ge et al. [34] researched a positive region-based reduction algorithm from the relative discernibility perspective in rough set theory. Sun and Xu [35] proposed a positive region-based granular space for feature selection based on rough sets. Nevertheless, the positive region in these models only draws attention to the consistent samples whose similarity classes are completely contained in some decision classes [36]. Meng et al. [37] presented an intersection neighborhood for numerical data, and designed a gene selection algorithm using positive region and gene ranking based on neighborhood rough sets. Liu et al. [38] designed a hash-based algorithm to calculate the positive region of neighborhood rough sets for attribute reduction. Li et al. [39] investigated a positive region-based related attribute reduction in neighborhood-based decision-theoretic rough set model. In summary, these literatures about attribute reduction are all based on rough sets or neighborhood rough sets from algebra view. And some of the abovementioned algorithms can achieve optimal reducts with criterion preservation, but in a sense, their reducts still have some redundant attributes that can be further deleted [2]. What's more, a lot of reduction algorithms still have higher time expenses when dealing with high-volume and high-dimensional data sets.

Until now, the attribute reduction models based on information theory have been studied extensively. Information entropy introduced by Shannon [40] is a very useful method for the representation of information contents in various domains [41]. Liang et al. [42] proposed information entropy, rough entropy and knowledge granulation for classifications in an incomplete information system. Sun et al. [43] presented some rough entropy-based uncertainty measures to improve computational efficiency of a heuristic feature selection algorithm. Gao et al. [2] introduced an attribute reduction algorithm based on the maximum decision entropy in decision-theoretic rough set model. Nevertheless, most of traditional rough sets-based methods for attribute reduction in the information view are not suitable to measure neighborhood classes of a real-value data set [20]. Liu et al. [44] combined information entropy with neighborhood rough sets to develop neighborhood mutual information. Chen et al. [20] constructed a gene selection method based on neighborhood rough sets and neighborhood entropy gain measures. Wang et al. [45] presented a feature selection method based on discrimination measures using conditional discrimination index in neighborhood rough sets. However, it is noted that the monotonicity of the abovementioned uncertainty measures does not always hold. In general, the attribute significance measures constructed in these algorithms can be applicable for numeric data sets, and these literatures for attribute reduction are based on rough sets or neighborhood rough sets from the information view.

As we can see, many existing methods of attribute reduction in neighborhood rough sets usually only start from the algebraic point of view or the information point of view, while the definition of attribute significance based on algebraic view only describes the effect of attributes on the subset of classification contained [46]. The definition of attribute significance based on information view only describes the influence of attributes on the uncertain classification subset contained in the domain and suitable for small-scale data sets [47]. Thus, they each have certain limitations in the real-world application. To overcome the shortcoming, considering to efficiently combine the above two views, Wang et al. [46] studied rough reduction in both the algebra view and the information view, and illustrated the definition of reduction and relative reduct in both the algebra view and the information view. Attribute reduction algorithms under the algebra and information viewpoints in rough set theory have been enhanced by filtering out redundant objects by Qian et al. [48]. Although these methods have their own advantages, they are still inefficient and not suitable for reducing large-scale high-dimensional data, and the enhanced algorithms only decrease the computation time to a certain

extent [47]. Inspired by this, to study neighborhood rough sets from the two views and achieve great uncertainty measures in neighborhood decision systems, the algebra view and the information view will be combined to develop attribute reduction algorithm for infinite sets in continuous-valued data sets. It follows that the Lebesgue measure [25] is introduced into neighborhood entropy to investigate the uncertainty measures in neighborhood decision systems, an attribute reduction method using Lebesgue and entropy measures is presented, and then a heuristic search algorithm is designed to analyze the uncertainty and noisy of continuous and complex data sets.

The rest of this paper is structured as follows: Some related concepts are briefly reviewed in Section 2. Section 3 describes Lebesgue measure-based neighborhood uncertainty measures, neighborhood entropy-based uncertainty measures, and comparison analysis with two representative reducts. Then, an attribute reduction algorithm with complexity analysis is presented. In Section 4, the classification experiments are conducted on public five UCI data sets and four gene expression data sets. Finally, Section 5 summarizes this study.

2. Previous Knowledge

In this section, we introduce some basic concepts and properties of decision system. The detailed descriptions can be found in literatures [22,41,49,50].

2.1. Rough Sets

Given a decision system $DS = \langle U, C, D, V, f \rangle$, usually written more simply as $DS = \langle U, C, D \rangle$, where $U = \{x_1, x_2, \dots, x_l\}$ is a sample set named universe; $C = \{a_1, a_2, \dots, a_s\}$ is a conditional attributes set that describes the samples; D is a set of classification decision attributes; $V = \cup_{a \in C \cup D} V_a$ and V_a is a value set of attribute a ; $f: U \times \{C \cup D\} \rightarrow V$ is a map function; and $f(a, x)$ represents the value of x on attribute $a \in C \cup D$.

Given a decision system $DS = \langle U, C, D \rangle$ with $B \subseteq C$, for any two samples $x, y \in U$, the equivalence relation [49] is described as:

$$IND(B) = \{(x, y) | \forall a \in B, f(a, x) = f(a, y)\} \quad (1)$$

Then, for any sample $x \in U$, $[x]_B = \{y | y \in U, (x, y) \in IND(B)\}$ is an equivalence class of x , and $U/IND(B)$ (U/B for short) is called as a partition that is composed of the equivalence classes.

The equivalence class defines two classical sets, named upper and lower approximation sets, as the elementary units [49]. Given a decision system $DS = \langle U, C, D \rangle$ with $B \subseteq C$ and $X \subseteq U$, the upper approximation set and the lower approximation set of X with respect to B can be described, respectively, as:

$$\bar{B}(X) = \{x | [x]_B \cap X \neq \emptyset, x \in U\}, \quad (2)$$

$$\underline{B}(X) = \{x | [x]_B \subseteq X, x \in U\}. \quad (3)$$

The precision is the ratio of the lower approximation set and the upper approximation set to measure the imprecision of a rough set. The roughness is an inverse of precision by a subtraction in the following.

Given a decision system $DS = \langle U, C, D \rangle$ with $B \subseteq C$ and $X \subseteq U$, the approximation precision of X with respect to B is described as:

$$\rho_B(X) = \frac{|\underline{B}(X)|}{|\bar{B}(X)|}. \quad (4)$$

Given a decision system $DS = \langle U, C, D \rangle$ with $B \subseteq C$ and $X \subseteq U$, the approximation roughness of X with respect to B is described as:

$$\gamma_B(X) = 1 - \rho_B(X) = 1 - \frac{|\underline{B}(X)|}{|\bar{B}(X)|}. \quad (5)$$

The approximation precision and the approximation roughness are used to measure the uncertainty and evaluate a rough set of information systems [41].

2.2. Information Entropy Measures

Given a decision system $DS = \langle U, C, D \rangle$ with $B \subseteq C$ and $U/B = \{X_1, X_2, \dots, X_n\}$, the information entropy [50] of B is described as:

$$H(B) = - \sum_{i=1}^n p(X_i) \log p(X_i), \tag{6}$$

where $p(X_i) = \frac{|X_i|}{|U|}$ is the probability of $X_i \subseteq U/B$, and $|X_i|$ denotes the cardinality of the equivalence class X_i .

Given a decision system $DS = \langle U, C, D \rangle$ with $B_1, B_2 \subseteq C$, $U/B_1 = \{X_1, X_2, \dots, X_n\}$, and $U/B_2 = \{Y_1, Y_2, \dots, Y_m\}$, then the joint entropy [50] of B_1 and B_2 is defined as:

$$H(B_1 \cup B_2) = - \sum_{i=1}^n \sum_{j=1}^m p(X_i \cap Y_j) \log p(X_i \cap Y_j), \tag{7}$$

where $p(X_i \cap Y_j) = \frac{|X_i \cap Y_j|}{|U|}$, $i = 1, 2, \dots, n$, and $j = 1, 2, \dots, m$.

Given a decision system $DS = \langle U, C, D \rangle$ with $B_1, B_2 \subseteq C$, $U/B_1 = \{X_1, X_2, \dots, X_n\}$, and $U/B_2 = \{Y_1, Y_2, \dots, Y_m\}$, then the conditional information entropy [50] of B_2 with respect to B_1 is defined as:

$$H(B_2|B_1) = - \sum_{i=1}^n p(X_i) \sum_{j=1}^m p(Y_j|X_i) \log p(Y_j|X_i), \tag{8}$$

where $p(Y_j|X_i) = \frac{|Y_j \cap X_i|}{|X_i|}$, $i = 1, 2, \dots, n$, and $j = 1, 2, \dots, m$.

2.3. Neighborhood Rough Sets

Given a real-value data set, which is formalized as a neighborhood decision system $NDS = \langle U, C, D, V, f, \Delta, \delta \rangle$, where $U = \{x_1, x_2, \dots, x_l\}$ is a sample set named universe; $C = \{a_1, a_2, \dots, a_s\}$ is a set of all conditional attributes; D is a decision attributes set; $V = \cup_{a \in C \cup D} V_a$ and V_a is a value set of attribute a ; $f: U \times \{C \cup D\} \rightarrow V$ is a map function; $\Delta \rightarrow [0, \infty]$ is a distance function; and δ is a neighborhood parameter with $0 \leq \delta \leq 1$. In the following, $NDS = \langle U, C, D, V, f, \Delta, \delta \rangle$ is simply noted as $NDS = \langle U, C \cup D, \delta \rangle$.

Given a neighborhood decision system $NDS = \langle U, C \cup D, \delta \rangle$ with $B \subseteq C$, a distance function $\Delta \rightarrow [0, \infty]$, and a neighborhood parameter $\delta \in [0, 1]$, then the neighborhood relation [22] is described as:

$$NR_\delta(B) = \{(x, y) \in U \times U | \Delta_B(x, y) \leq \delta\}. \tag{9}$$

According to the definition of neighborhood relation, for any $x \in U$, the neighborhood class of x with respect to $B \subseteq C$ is expressed as:

$$n_B^\delta(x) = \{y | x, y \in U, \Delta_B(x, y) \leq \delta\}. \tag{10}$$

Because the Euclidean distance function effectively reflects the basic information of the unknown data [22], it is introduced into this paper, and its formula is expressed as:

$$\Delta_B(x, y) = \sqrt{\sum_{k=1}^N |f(a_k, x) - f(a_k, y)|^2}, \tag{11}$$

where N is the cardinality of subset B .

Given a neighborhood decision system $NDS = \langle U, C \cup D, \delta \rangle$ with $B \subseteq C$ and $X \subseteq U$, the neighborhood upper approximation set and the neighborhood lower approximation set of X with respect to B are denoted, respectively, as:

$$\overline{B}(X)_\delta = \{x_i | n_B^\delta(x_i) \cap X \neq \emptyset, x_i \in U\}, \tag{12}$$

$$\underline{B}(X)_\delta = \{x_i | n_B^\delta(x_i) \subseteq X, x_i \in U\}. \tag{13}$$

3. Attribute Reduction Using Lebesgue and Entropy Measures in Neighborhood Decision Systems

3.1. Lebesgue Measure-Based Neighborhood Uncertainty Measures

Aiming at the problem that the existing neighborhood rough set model cannot handle the infinite sets, a neighborhood rough set model combined with Lebesgue measure is proposed, which is on the basis of the neighborhood rough set model and the measure theory. Then, the concept of Lebesgue measure is introduced to extend neighborhood rough sets for an infinite set.

For any M -dimensional Euclidean space R^M , let E be a point set in R^M , and for an open interval I_i of each column covered E , $\bigcup_{i=1}^\infty I_i \supset E$ holds. Then, the sum of its volume is $\mu = \sum_{i=1}^\infty |I_i|$, and all of μ form a bounded below set of numbers. The infimum is called the Lebesgue outer measure of E , denoted as $m^*(E)$, i.e.:

$$m^*(E) = \inf_{E \subseteq \bigcup_{i=1}^\infty I_i} \sum_{i=1}^\infty |I_i|. \tag{14}$$

The Lebesgue inner measure can be described as $m_*(E) = |I| - m^*(I - E)$. If $m_*(E) = m^*(E)$, then E is said to be measurable, denoted as the Lebesgue measure $m(E)$. When the Lebesgue measure of U is 0, it can be shown as the cardinality of U , i.e., $|U|$. Here, $m(X)$ is used uniformly to describe the Lebesgue measure of a set X in this paper.

Definition 1. Given a neighborhood decision system $NDS = \langle U, C \cup D, \delta \rangle$ with non-empty infinite set U and any $B \subseteq C$, $\Delta_B(x, y)$ is a distance function between two objects, and a neighborhood parameter $0 \leq \delta \leq 1$, for any $x, y \in U$, then a Lebesgue measure of neighborhood class with respect to B is defined as

$$m(n_B^\delta(x)) = m(\{y | x, y \in U, \Delta_B(x, y) \leq \delta\}). \tag{15}$$

Proposition 1. Given a neighborhood decision system $NDS = \langle U, C \cup D, \delta \rangle$ with non-empty infinite set U , and $P, Q \subseteq C$, for any $x \in U$, then the following properties hold:

- (1) $m(U) = |U|$.
- (2) If $Q \subseteq P$, then $m(n_P^\delta(x)) \leq m(n_Q^\delta(x))$.
- (3) If $0 \leq \alpha \leq \delta \leq 1$, then $m(n_P^\alpha(x)) \leq m(n_P^\delta(x))$.
- (4) For any $q \in Q$, $m(n_Q^\delta(x)) \leq m(\bigcap_{q \in Q} n_q^\delta(x))$.
- (5) $m(n_P^\delta(x)) \neq 0$ and $m(\bigcup_{x \in U} n_P^\delta(x)) = m(U)$.

Proof.

- (1) This proof is straightforward.
- (2) Suppose that any attribute subset $Q, R \subseteq P \subseteq C$, there must exist $P = Q \cup R$. From Equation (10), one has that $n_Q^\delta(x) = \{y | x, y \in U, \Delta_Q(x, y) \leq \delta\}$ and $n_{Q \cup R}^\delta(x) = \{y | x, y \in U, \Delta_{Q \cup R}(x, y) \leq \delta\}$. Then, it can

be obtained from Proposition 1 in [20] that $\Delta_Q(x, y) \leq \Delta_{Q \cup R}(x, y)$, i.e., $\Delta_Q(x, y) \leq \Delta_P(x, y)$. It follows from Equation (10) that $n_P^\delta(x) \subseteq n_Q^\delta(x)$. Therefore, by Equation (15), $m(n_P^\delta(x)) \leq m(n_Q^\delta(x))$ holds.

(3) For any $0 \leq \alpha \leq \delta \leq 1$, it follows immediately from Proposition 1 in [41] that $n_P^\alpha(x) \subseteq n_P^\delta(x)$ holds. Hence, one has $m(n_P^\alpha(x)) \leq m(n_P^\delta(x))$.

(4) Suppose that any $q \in Q \subseteq C$, it follows from Proposition 1 in [41] that $n_Q^\delta(x) \subseteq \bigcap_{q \in Q} n_q^\delta(x)$. Then, one has that $m(n_Q^\delta(x)) \leq m(\bigcap_{q \in Q} n_q^\delta(x))$.

(5) For any $P \subseteq C$, it can be obtained from Proposition 1 in [41] that $n_P^\delta(x) \neq \emptyset$ and $\bigcup_{x \in U} n_P^\delta(x) = U$. Hence, both $m(n_P^\delta(x)) \neq 0$ and $m(\bigcup_{x \in U} n_P^\delta(x)) = m(U)$ hold. \square

Definition 2. Given a neighborhood decision system $NDS = \langle U, C \cup D, \delta \rangle$ with non-empty infinite set U , $B \subseteq C$ and $X \subseteq U$, an upper approximation and a lower approximation of X with respect to B based on Lebesgue measure are defined, respectively, as:

$$m(\overline{B}(X)_\delta) = m(\{x | n_B^\delta(x) \cap X \neq \emptyset, x \in U\}), \tag{16}$$

$$m(\underline{B}(X)_\delta) = m(\{x | n_B^\delta(x) \subseteq X, x \in U\}). \tag{17}$$

Proposition 2. Given a neighborhood decision system $NDS = \langle U, C \cup D, \delta \rangle$ with non-empty infinite set U , $B \subseteq C$ and $X, Y \subseteq U$, then the following properties hold:

- (1) $m(\underline{B}(X)_\delta) \leq m(\overline{B}(X)_\delta)$.
- (2) $m(\overline{B}(X \cup Y)_\delta) = m(\overline{B}(X)_\delta) + m(\overline{B}(Y)_\delta)$.
- (3) $m(\underline{B}(X \cup Y)_\delta) \geq m(\underline{B}(X)_\delta) + m(\underline{B}(Y)_\delta)$.
- (4) $X \subseteq Y \Rightarrow m(\underline{B}(X)_\delta) \leq m(\underline{B}(Y)_\delta)$.
- (5) $X \subseteq Y \Rightarrow m(\overline{B}(X)_\delta) \leq m(\overline{B}(Y)_\delta)$.

Proof.

(1) This proof is straightforward.

(2) Suppose that any $x \in \overline{B}(X \cup Y)_\delta$, it follows from Equation (12) that $n_B^\delta(x) \cap (X \cup Y) \neq \emptyset$, and then $(n_B^\delta(x) \cap X) \cup (n_B^\delta(x) \cap Y) \neq \emptyset$. It is obvious that $n_B^\delta(x) \cap X \neq \emptyset$ or $n_B^\delta(x) \cap Y \neq \emptyset$. From Equation (12), one has that $x \in \overline{B}(X)_\delta$ or $x \in \overline{B}(Y)_\delta$, and then $x \in \overline{B}(X)_\delta \cup \overline{B}(Y)_\delta$. Thus, it can be obtained that $\overline{B}(X \cup Y)_\delta = \overline{B}(X)_\delta \cup \overline{B}(Y)_\delta$. Therefore, one has $m(\overline{B}(X \cup Y)_\delta) = m(\overline{B}(X)_\delta) + m(\overline{B}(Y)_\delta)$.

(3) Since there exist $X \subseteq X \cup Y \subseteq U$ and $Y \subseteq X \cup Y \subseteq U$, and it follows from Equation (5) in [49] that $\underline{B}(X)_\delta \subseteq \underline{B}(X \cup Y)_\delta$ and $\underline{B}(Y)_\delta \subseteq \underline{B}(X \cup Y)_\delta$, which yields $\underline{B}(X)_\delta \cup \underline{B}(Y)_\delta \subseteq \underline{B}(X \cup Y)_\delta$. Obviously, $m(\underline{B}(X)_\delta \cup \underline{B}(Y)_\delta) \leq m(\underline{B}(X \cup Y)_\delta)$ such that one has $m(\underline{B}(X)_\delta) + m(\underline{B}(Y)_\delta) \leq m(\underline{B}(X \cup Y)_\delta)$. Hence, $m(\underline{B}(X \cup Y)_\delta) \geq m(\underline{B}(X)_\delta) + m(\underline{B}(Y)_\delta)$ can be obtained.

(4) Suppose that $X \subseteq Y$, it follows that $X \cap Y = X$, and then $\underline{B}(X \cap Y) = \underline{B}(X)$. Similar to the Equation (5) in [49], it can be obtained that $\underline{B}(X) \cap \underline{B}(Y) = \underline{B}(X)$ in rough sets. Obviously, $\underline{B}(X \cap Y)_\delta = \underline{B}(X)_\delta$, and then $\underline{B}(X)_\delta \cap \underline{B}(Y)_\delta = \underline{B}(X)_\delta$. One has that $\underline{B}(X)_\delta \subseteq \underline{B}(Y)_\delta$. Hence, $m(\underline{B}(X)_\delta) \leq m(\underline{B}(Y)_\delta)$ holds.

(5) For any $X \subseteq Y$, it follows that $X \cup Y = Y$, and then $\overline{B}(X \cup Y)_\delta = \overline{B}(Y)_\delta$. From the proof of (2), one has that $\overline{B}(X)_\delta \cup \overline{B}(Y)_\delta = \overline{B}(Y)_\delta$. Obviously, $\overline{B}(X)_\delta \subseteq \overline{B}(Y)_\delta$. Thus, $m(\overline{B}(X)_\delta) \leq m(\overline{B}(Y)_\delta)$ holds. \square

Definition 3. Given a neighborhood decision system $NDS = \langle U, C \cup D, \delta \rangle$ with non-empty infinite set U , $B \subseteq C$ and $X \subseteq U$, a new neighborhood approximate precision of X with respect to B based on Lebesgue measure is defined as:

$$\rho_B^\delta(X) = \frac{m(\underline{B}(X)_\delta)}{m(\overline{B}(X)_\delta)}. \tag{18}$$

In a neighborhood decision system, the measure of neighborhood approximate precision is the percentage of possible correct decisions when classifying objects. It is monotonically increasing with the growth of a conditional attribute.

Property 1. Given a neighborhood decision system $NDS = \langle U, C \cup D, \delta \rangle$ with non-empty infinite set U , $B \subseteq C$ and $X \subseteq U$, then there exists $0 \leq \rho_B^\delta(X) \leq 1$.

Proof. Suppose that $B \subseteq C$ and $X \subseteq U$, it follows from Proposition 2 that $m(\underline{B}(X)_\delta) \leq m(\overline{B}(X)_\delta)$ established. Then, it is obvious that $0 \leq \frac{m(\underline{B}(X)_\delta)}{m(\overline{B}(X)_\delta)} \leq 1$, and from Equation (18), one has that $0 \leq \rho_B^\delta(X) \leq 1$. \square

The neighborhood approximation precision $\rho_B^\delta(X)$ is used to reflect the degree of knowledge of acquiring set X . When $\rho_B^\delta(X) = 1$, the B boundary of X is an empty set. At this time, the set X is precisely defined on B . When $\rho_B^\delta(X) < 1$, the set X has a non-empty B boundary domain, and the set X is undefined on B . Of course, some other metrics can also be used to define the imprecision of X .

Definition 4. Given a neighborhood decision system $NDS = \langle U, C \cup D, \delta \rangle$ with non-empty infinite set U , $B \subseteq C$ and $X \subseteq U$, a new neighborhood roughness of X with respect to B based on Lebesgue measure is defined as:

$$\gamma_B^\delta(X) = 1 - \rho_B^\delta(X) = 1 - \frac{m(\underline{B}(X)_\delta)}{m(\overline{B}(X)_\delta)}. \quad (19)$$

The neighborhood roughness of X with respect to B is opposite and complementary to the neighborhood approximate precision. It represents the degree of incompleteness of obtaining knowledge of set X .

Property 2. Given a neighborhood decision system $NDS = \langle U, C \cup D, \delta \rangle$ with non-empty infinite set U , $B \subseteq C$ and $X \subseteq U$, then there exists $0 \leq \gamma_B^\delta(X) \leq 1$.

Proof. It follows immediately from Property 1 that the proof is straightforward. \square

Proposition 3. Given a neighborhood decision system $NDS = \langle U, C \cup D, \delta \rangle$ with non-empty infinite set U , $Q \subseteq P \subseteq C$ and $X \subseteq U$, then one has $\rho_Q^\delta(X) \leq \rho_P^\delta(X)$ and $\gamma_Q^\delta(X) \geq \gamma_P^\delta(X)$.

Proof. Suppose that for any $Q \subseteq P \subseteq C$ and $x \in U$, it follows from Proposition 1 that $m(n_P^\delta(x)) \leq m(n_Q^\delta(x))$. By Equation (12), $\overline{Q}(X)_\delta = \{x | n_Q^\delta(x) \cap X \neq \emptyset, x \in U\}$ and $\overline{P}(X)_\delta = \{x | n_P^\delta(x) \cap X \neq \emptyset, x \in U\}$ hold. Similarly, from Equation (13), one has $\underline{Q}(X)_\delta = \{x | n_Q^\delta(x) \subseteq X, x \in U\}$ and $\underline{P}(X)_\delta = \{x | n_P^\delta(x) \subseteq X, x \in U\}$. It can be obtained that $\underline{Q}(X)_\delta \subseteq \underline{P}(X)_\delta$ and $\overline{Q}(X)_\delta \subseteq \overline{P}(X)_\delta$. It follows from Proposition 2 that $m(\underline{Q}(X)_\delta) \leq m(\underline{P}(X)_\delta)$ and $m(\overline{Q}(X)_\delta) \leq m(\overline{P}(X)_\delta)$. Then, it is obvious that $\frac{m(\underline{Q}(X)_\delta)}{m(\overline{Q}(X)_\delta)} \leq \frac{m(\underline{P}(X)_\delta)}{m(\overline{P}(X)_\delta)}$ so that one has $1 - \frac{m(\underline{Q}(X)_\delta)}{m(\overline{Q}(X)_\delta)} \geq 1 - \frac{m(\underline{P}(X)_\delta)}{m(\overline{P}(X)_\delta)}$. Therefore, both $\rho_Q^\delta(X) \leq \rho_P^\delta(X)$ and $\gamma_Q^\delta(X) \geq \gamma_P^\delta(X)$ hold. \square

It is known that the classical measurement methods are used to estimate a set of data classified in a knowledge system [3]. From Proposition 3, the neighborhood approximate precision and the neighborhood roughness are used to measure the uncertainty of rough classification in neighborhood decision systems.

3.2. Neighborhood Entropy-Based Uncertainty Measures

In rough set theory, information entropy is used as a measure to assess the value of equivalence class in a discrete decision system [50]. However, it is not appropriate to measure the neighborhood classes in real-value data sets [18]. To solve this problem, the concept of neighborhood has been

introduced into information entropy to extend Shannon entropy [51]. But, most of neighborhood entropy-based measures and their variations only analyze finite sets, which would limit the practical application of neighborhood rough sets to a certain degree. It is known that the Lebesgue measure can measure the uncertainty of the infinite sets [25]. Then, the Lebesgue measure is introduced to study the uncertainty measures of infinite sets in neighborhood decision systems.

Given a neighborhood decision system $NDS = \langle U, C \cup D, \delta \rangle$ with $B \subseteq C$, $n_B^\delta(x_i)$ is a neighborhood class of $x_i \in U$, and then Hu et al. [21] described the neighborhood entropy of x_i with respect to B as follows:

$$H_\delta^{x_i}(B) = -\log \frac{|n_B^\delta(x_i)|}{|U|}. \tag{20}$$

Definition 5. Given a neighborhood decision system $NDS = \langle U, C \cup D, \delta \rangle$ with non-empty infinite set U , $B \subseteq C$ and $x_i \in U$, a new neighborhood entropy of B based on Lebesgue measure is defined as:

$$H_\delta^{x_i}(B) = -\log_2 \frac{m(n_B^\delta(x_i))}{m(U)}. \tag{21}$$

Definition 6. Given a neighborhood decision system $NDS = \langle U, C \cup D, \delta \rangle$ with non-empty infinite set U , and $B \subseteq C$, an average neighborhood entropy of the universe U based on Lebesgue measure is defined as:

$$H_\delta(B) = -\frac{1}{m(U)} \sum_{i=1}^{|U|} \log_2 \frac{m(n_B^\delta(x_i))}{m(U)}. \tag{22}$$

Proposition 4. Given a neighborhood decision system $NDS = \langle U, C \cup D, \delta \rangle$ with non-empty infinite set U and $P, Q \subseteq C$, for any $x_i \in U$, if $n_P^\delta(x_i) = n_Q^\delta(x_i)$, then $H_\delta(P) = H_\delta(Q)$.

Proof. Suppose that for any $x_i \in U$, $n_P^\delta(x_i) = n_Q^\delta(x_i)$, and it follows that from Equation (15) that $m(n_P^\delta(x_i)) = m(n_Q^\delta(x_i))$. Then, it can be obtained that $\sum_{i=1}^{|U|} \log_2 \frac{m(n_P^\delta(x_i))}{m(U)} = \sum_{i=1}^{|U|} \log_2 \frac{m(n_Q^\delta(x_i))}{m(U)}$. Hence, by Equation (22), one has $H_\delta(P) = H_\delta(Q)$. \square

Definition 7. Given a neighborhood decision system $NDS = \langle U, C \cup D, \delta \rangle$ with non-empty infinite set U and $B \subseteq C$, $n_B^\delta(x_i)$ is the neighborhood class of x_i in neighborhood relation, and $[x_i]_d$ is an equivalence class formed by the decision attribute d of x_i in equivalence relation. Then, a joint entropy of subsets B and d based on Lebesgue measure is defined as:

$$H_\delta(Bd) = -\frac{1}{m(U)} \sum_{i=1}^{|U|} \log_2 \left(\frac{m(n_B^\delta(x_i) \cap [x_i]_d)}{m(U)} \right). \tag{23}$$

Proposition 5. Given a neighborhood decision system $NDS = \langle U, C \cup \{d\}, \delta \rangle$ with non-empty infinite set U , and $Q \subseteq P \subseteq C$, then $H_\delta(Qd) \leq H_\delta(Pd)$.

Proof Suppose that any $Q \subseteq P \subseteq C$, it can be obtained from Proposition 1 that $n_P^\delta(x) \subseteq n_Q^\delta(x)$ and $m(n_P^\delta(x)) \leq m(n_Q^\delta(x))$. Let $[x_i]_d$ be an equivalence class formed by d of x_i in equivalence relation. Then, $\{x_i\} \subseteq n_P^\delta(x_i) \cap [x_i]_d \subseteq n_Q^\delta(x_i) \cap [x_i]_d \subseteq U$ holds. Obviously, one has that $m(\{x_i\}) \leq m(n_P^\delta(x_i) \cap [x_i]_d) \leq m(n_Q^\delta(x_i) \cap [x_i]_d) \leq m(U)$. It follows that $\frac{1}{m(U)} \leq \frac{m(n_P^\delta(x_i) \cap [x_i]_d)}{m(U)} \leq \frac{m(n_Q^\delta(x_i) \cap [x_i]_d)}{m(U)} \leq 1$. Then, it is obvious that $\log_2\left(\frac{1}{m(U)}\right) \leq \log_2\left(\frac{m(n_P^\delta(x_i) \cap [x_i]_d)}{m(U)}\right) \leq \log_2\left(\frac{m(n_Q^\delta(x_i) \cap [x_i]_d)}{m(U)}\right) \leq 0$. Thus, it can

be obtained that $0 \leq -\frac{1}{m(U)} \sum_{i=1}^{|U|} \log_2 \left(\frac{m(n_B^\delta(x_i) \cap [x_i]_d)}{m(U)} \right) \leq -\frac{1}{m(U)} \sum_{i=1}^{|U|} \log_2 \left(\frac{m(n_P^\delta(x_i) \cap [x_i]_d)}{m(U)} \right) \leq \log_2 m(U)$.
 Therefore, $H_\delta(Qd) \leq H_\delta(Pd)$ holds. \square

Definition 8. Given a neighborhood decision system $NDS = \langle U, C \cup \{d\}, \delta \rangle$ with non-empty infinite set U and $B \subseteq C$, for any class of object $x \in U$ with respect to B , $n_B^\delta(x)$ is a neighborhood class of x generated by the neighborhood relation $NR_\delta(B)$, $[x_i]_d$ is an equivalence class of x_i generated by equivalence relation $IND(d)$, and $U/d = \{d_1, d_2, \dots, d_t, \dots\}$. Then, the neighborhood roughness joint entropy based on Lebesgue measure of d with respect to B is defined as:

$$NRH(d, B) = -\frac{1}{m(U)} \sum_{j=1}^{\infty} \log_2 (2^{-\gamma_B^\delta(d_j)}) \times \int_{x_i \in U} \log_2 \left(\frac{m(n_B^\delta(x_i) \cap d_j)}{m(U)} \right) dx. \tag{24}$$

It is noted that Wang et al. [46] stated that all conceptions and computations in rough set theory based on the upper and lower approximation sets are called the algebra view of the rough set theory, and the notions of information entropy and its extensions are called the information view of rough sets. It follows from Equation (24) that $\gamma_B^\delta(d_i)$ is the neighborhood roughness of d_i with respect to B in the algebra view and it represents the degree of incompleteness of obtaining knowledge of set d_i , and $\sum_{j=1}^{\infty} \int_{x_i \in U} \log_2 \left(\frac{m(n_B^\delta(x_i) \cap d_j)}{m(U)} \right) dx$ is the definition of joint entropy in the information view. Hence, Definition 8 can efficiently analyze and measure the uncertainty of neighborhood decision systems based on Lebesgue and entropy measures from both the algebra view and the information view.

Property 3. Given a neighborhood decision system $NDS = \langle U, C \cup \{d\}, \delta \rangle$ with non-empty infinite set U , $B \subseteq C$, and $U/d = \{d_1, d_2, \dots, d_t, \dots\}$, then

$$NRH(d, B) = -\frac{1}{|U|} \sum_{j=1}^{\infty} \log_2 (2^{-\gamma_B^\delta(d_j)}) \times \int_{x_i \in U} \log_2 \left(\frac{m(n_B^\delta(x_i) \cap d_j)}{|U|} \right) dx \geq 0$$

Proof. Suppose that for any $B \subseteq C$, it follows from Proposition 1 that $m(U) = |U|$. Then, one has that $NRH(d, B) = -\frac{1}{|U|} \sum_{j=1}^{\infty} \log_2 (2^{-\gamma_B^\delta(d_j)}) \times \int_{x_i \in U} \log_2 \left(\frac{m(n_B^\delta(x_i) \cap d_j)}{|U|} \right) dx$. As known from Equation (23), due to $m(n_B^\delta(x_i) \cap d_j) \leq m(U)$, it is obvious that $\log_2 \left(\frac{m(n_B^\delta(x_i) \cap d_j)}{|U|} \right) \leq 0$ and $\log_2 (2^{-\gamma_B^\delta(d_j)}) \geq 0$. Therefore, $NRH(d, B) \geq 0$ holds. \square

Proposition 6. Given a neighborhood decision system $NDS = \langle U, C \cup D, \delta \rangle$ with non-empty infinite set U , $P \subseteq Q \subseteq C$, and $U/d = \{d_1, d_2, \dots, d_t, \dots\}$, then $NRH(D, Q) \leq NRH(D, P)$.

Proof. Suppose that for any $Q \subseteq P \subseteq C$ and $X \subseteq U$, it follows from Proposition 3 that $\rho_Q^\delta(X) \leq \rho_P^\delta(X)$ and $\gamma_P^\delta(X) \leq \gamma_Q^\delta(X)$. It is clear that $0 \leq \log (2^{-\gamma_Q^\delta(d_i)}) \leq \log (2^{-\gamma_P^\delta(d_i)})$. From Proposition 5, one has that $H_\delta(Qd) \leq H_\delta(Pd)$. Then, it can be obviously obtained that $-\frac{1}{m(U)} \int_{x_i \in U} \log_2 \left(\frac{m(n_Q^\delta(x_i) \cap d_j)}{|U|} \right) dx \leq -\frac{1}{m(U)} \int_{x_i \in U} \log_2 \left(\frac{m(n_P^\delta(x_i) \cap d_j)}{|U|} \right) dx$. When $n_Q^\delta(x) = n_P^\delta(x)$ for any $x \in U$, one has $\gamma_P^\delta(d_j) = \gamma_Q^\delta(d_j)$ and $n_Q^\delta(x) \cap d_j = n_P^\delta(x) \cap d_j$, where $1 \leq j$. Thus, $NRH(D, Q) = NRH(D, P)$. Therefore, $NRH(D, Q) \leq NRH(D, P)$ holds. \square

The monotonicity is one of the most important properties for an effective uncertainty measure of attribute reduction. According to Proposition 6, it is quite obvious that the neighborhood roughness

joint entropy is monotonically increasing when adding the conditional attributes, which validates the monotonicity of the proposed uncertainty measure. Furthermore, the monotonicity contributes to the selection of the greedy method for attribute reduction.

Definition 9. Given a neighborhood decision system $NDS = \langle U, C \cup D, \delta \rangle$ with non-empty infinite set U and $B \subseteq C$, for any $a \in B$, then the internal attribute significance of a in B relative to D is defined as:

$$\text{Sig}^{\text{inner}}(a, B, D) = \text{NRH}(D, B) - \text{NRH}(D, B - \{a\}). \quad (25)$$

Definition 10. Given a neighborhood decision system $NDS = \langle U, C \cup D, \delta \rangle$ with non-empty infinite set U and $B \subseteq C$, for any $a \in C - B$, then the external attribute significance of a relative to D is defined as:

$$\text{Sig}^{\text{outer}}(a, B, D) = \text{NRH}(D, B \cup \{a\}) - \text{NRH}(D, B) \quad (26)$$

Definition 11. Given a neighborhood decision system $NDS = \langle U, C \cup D, \delta \rangle$ with non-empty infinite set U and any $a \in C$, if $\text{NRH}(D, C) > \text{NRH}(D, C - \{a\})$, that is, $\text{Sig}^{\text{inner}}(a, C, D) > 0$, then the attribute a is a core of C relative to D .

Definition 12. Given a neighborhood decision system $NDS = \langle U, C \cup D, \delta \rangle$ with non-empty infinite set U and $B \subseteq C$, if any $a \in B$ is necessary in B if and only if $\text{Sig}^{\text{inner}}(a, B, D) > 0$; otherwise a is unnecessary. If each $a \in B$ is necessary, one can say that B is independent; otherwise B is dependent.

Definition 13. Given a neighborhood decision system $NDS = \langle U, C \cup D, \delta \rangle$ with non-empty infinite set U and $B \subseteq C$, if $\text{NRH}(D, B) = \text{NRH}(D, C)$, and for any $a \in B$, there exists $\text{NRH}(D, B) > \text{NRH}(D, B - \{a\})$, then it is said that B is a reduct of C relative to D .

3.3. Comparative Analysis with Two Representative Reducts

It is known that the definition of reducts from the algebra view is usually equivalent to its definition from the information view in a general information system. What's more, the relative reduct of a decision system in the information view includes that in the algebra view. Thus, Wang et al. [46] declared that any relative reduct of a decision system in the information view must be its relative reduct in the algebra view, so that some heuristic algorithms can be designed further using this conclusion. Based on the ideas of the classification in [46], the definition of reducts based on neighborhood roughness joint entropy of a neighborhood decision system should be developed from the algebra view and the information view in neighborhood rough set theory. For convenience, the reduct in Definition 13 is named as the neighborhood entropy reduct. Liu et al. [38] presented a reduct based on positive region in the neighborhood decision system similar with classical rough set model. This representative relative reduct based on positive region is called the algebra view of the neighborhood rough set theory. Chen et al. [20] defined information quantity similar to information entropy to evaluate the neighborhood classes, used the joint entropy gain to evaluate the significance of a selecting attribute, and proposed a representative joint entropy gain-based reduction algorithm, which is called a reduct in the information view of neighborhood rough sets.

Given a neighborhood decision system $NDS = \langle U, C \cup D, \delta \rangle$ with non-empty infinite set U , $B \subseteq C$ and $D = \{d\}$. Then, a positive region reduct of the neighborhood decision system is presented as follows in [38]: for any $a \in B$, if $|\text{POS}_B(D)| = |\text{POS}_C(D)|$ and $|\text{POS}_{B - \{a\}}(D)| < |\text{POS}_B(D)|$, where $\text{POS}_B(D) = \cup\{B(X)_\delta | X \in U/D\}$ is the positive region of D with respect to B , B is a relative reduct of the neighborhood decision system.

Proposition 7. *Given a neighborhood decision system $NDS = \langle U, C \cup D, \delta \rangle$ with non-empty infinite set U , and $B \subseteq C$, if B is a neighborhood entropy reduct of the neighborhood decision system, then B is a positive region reduct of the neighborhood decision system.*

Proof. Let $U = \{x_1, x_2, \dots, x_n, \dots\}$, and $U/D = \{d_1, d_2, \dots, d_t, \dots\}$. Suppose that for a subset $B \subseteq C$, it follows from Definition 13 that if $NRH(D, B) = NRH(D, C)$, and for any $a \in B$, there exists $NRH(D, B) > NRH(D, B - \{a\})$, then B is a neighborhood entropy reduct of C relative to D . When $NRH(D, B) = NRH(D, C)$, it can be obtained from Proposition 6 that $n_B^\delta(x) = n_C^\delta(x)$, $\gamma_B^\delta(d_j) = \gamma_C^\delta(d_j)$ and $n_B^\delta(x) \cap d_j = n_C^\delta(x) \cap d_j$ hold, where any $x \in U$ and $1 \leq j$. By Equation (13), one has that $\underline{B}(D)_\delta = \underline{C}(D)_\delta$. So it is obvious that $POS_B(D) = POS_C(D)$, i.e., $|POS_B(D)| = |POS_C(D)|$. For any $a \in B$, $B - \{a\} \subset B$, and from Theorem 1 in [38], one has that $\overline{B - \{a\}}(D)_\delta \subseteq \overline{B}(D)_\delta$, so that $POS_{B - \{a\}}(D) \subseteq POS_B(D)$ holds. Because for any $a \in B$, there exists $NRH(D, B) > NRH(D, B - \{a\})$, thus $\overline{B - \{a\}}(D)_\delta \subset \overline{B}(D)_\delta$ holds. It follows that $POS_{B - \{a\}}(D) \subset POS_B(D)$. Thus, $|POS_{B - \{a\}}(D)| < |POS_B(D)|$ for any $a \in B$. Therefore, B is a positive region reduct of the neighborhood decision system. \square

Notably, the inverse relation of this proposition generally does not hold. According to the above discussions, Proposition 7 shows that the definition of the neighborhood entropy reduct includes that of positive region reduct in the algebra view.

Given a neighborhood decision system $NDS = \langle U, C \cup D, \delta \rangle$ with non-empty infinite set U , $B \subseteq C$ and $D = \{d\}$. For any $a \in B$, a reduct of the neighborhood decision system, named as the entropy gain reduct is proposed in [20] as follows: if $H(Bd) = H(Cd)$ and $H(\{B - \{a\}\}d) < H(Bd)$, where $H(Bd) = -\frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \left(\frac{n_B^\delta(x_i) \cap [x_i]_d}{|U|} \right)$ describes the joint entropy of B and d , B is an entropy gain reduct of the neighborhood decision system.

Proposition 8. *Given a neighborhood decision system $NDS = \langle U, C \cup D, \delta \rangle$ with non-empty infinite set U and $B \subseteq C$, then B is a neighborhood entropy reduct of the neighborhood decision system if and only if B is an entropy gain reduct of the neighborhood decision system.*

Proof. Let $U = \{x_1, x_2, \dots, x_n, \dots\}$, and $U/D = \{d_1, d_2, \dots, d_t, \dots\}$. Suppose that for a subset $B \subseteq C$, it follows from Definition 13 that if $NRH(D, B) = NRH(D, C)$, and for any $a \in B$, there exists $NRH(D, B) > NRH(D, B - \{a\})$, then B is a neighborhood entropy reduct of C relative to D . Similar to the proof of Proposition 7, when $NRH(D, B) = NRH(D, C)$, from Proposition 6, one has that $n_B^\delta(x) \cap d_j = n_C^\delta(x) \cap d_j$, where any $x \in U$ and $1 \leq j$. It is obvious that $H(Bd) = H(Cd)$. Since $B - \{a\} \subset B$, from Proposition 2 in [20], one has that $H(\{B - \{a\}\}d) \leq H(Bd)$. Because for any $a \in B$, there exists $NRH(D, B) > NRH(D, B - \{a\})$, so $H(\{B - \{a\}\}d) < H(Bd)$ holds. Hence, B is an entropy gain reduct of the neighborhood decision system.

Suppose that for a subset $B \subseteq C$, and any $a \in B$, if $H(Bd) = H(Cd)$ and $H(\{B - \{a\}\}d) < H(Bd)$, then B is an entropy gain reduct of C relative to D . Similar to the proof of Proposition 6, when $n_B^\delta(x) = n_C^\delta(x)$, by Equations (16), (17) and (19), one has that $\gamma_B^\delta(d_j) = \gamma_C^\delta(d_j)$, and then it is obvious that $n_B^\delta(x) \cap d_j = n_C^\delta(x) \cap d_j$, where any $x \in U$ and $1 \leq j$. Thus, it can be obtained from Equation (24) that $NRH(D, B) = NRH(D, C)$. Because $B - \{a\} \subset B$, it follows from Proposition 6 that $NRH(D, B - \{a\}) \leq NRH(D, B)$. Since for any $a \in B$, there exists $H(\{B - \{a\}\}d) < H(Bd)$. So, one has that $NRH(D, B - \{a\}) < NRH(D, B)$. Therefore, B is a neighborhood entropy reduct of the neighborhood decision system. \square

Proposition 8 shows that in a neighborhood decision system, the neighborhood entropy reduct is equivalent to the entropy gain reduct in the information view. According to Propositions 7 and 8, it can be concluded that the definition of neighborhood entropy reduct includes two representative reducts proposed in the algebra view and the information view. Therefore, the definition of neighborhood entropy reduct denotes a mathematical quantitative measure to evaluate the knowledge uncertainty of different attribute sets in neighborhood decision systems.

3.4. Description of the Attribute Reduction Algorithm

In order to facilitate the understanding of the attribute reduction method, the process of attribute reduction algorithm for data classification is illustrated in Figure 1.

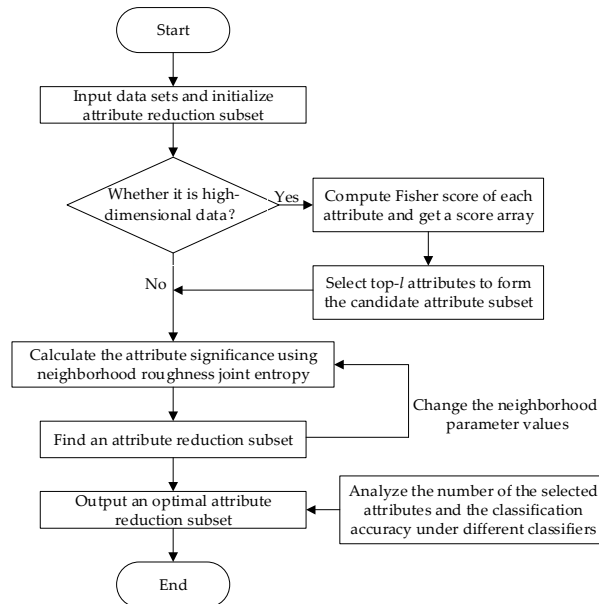


Figure 1. Flowchart of the attribute reduction algorithm for data classification.

To support efficient attribute reduction, an attribute reduction algorithm based on neighborhood roughness joint entropy (ARNRJE) is constructed and described as Algorithm 1.

Algorithm 1. ARNRJE

Input: A neighborhood decision system $NDS = \langle U, C \cup D, \delta \rangle$, and neighborhood parameter δ .

Output: An optimal reduction set B .

1. Initialize $B = \emptyset$.
 2. Calculate $NRH(D, C)$.
 3. For $i = 1$ to $|C|$ do // Obtain the attribute core
 4. Calculate $Sig^{inner}(a_i, C, D)$.
 5. If $Sig^{inner}(a_i, C, D) > 0$, then $B = B \cup \{a_i\}$.
 6. End for
 7. Let $R = C - B$.
 8. Do
 9. For $j = 1$ to $|R|$
 - a) Calculate $NRH(D, B \cup \{a_j\})$.
 - b) Select a_j to make it satisfy $\max\{a_j \in R \mid NRH(D, B \cup \{a_j\})\}$, and if there are multiple attributes that satisfy the maximum, then the front should be selected. End for
 10. Let $B = B \cup \{a_j\}$ and $R = R - \{a_j\}$, and calculate $NRH(D, B)$.
 11. While $NRH(D, B) \neq NRH(D, C)$
 12. For $k = 1$ to $|B|$ do // Verify the completeness of the reduction subset
 13. Select $a_k \in B$.
 14. Compute $NRH(D, B - \{a_k\})$.
 15. If $NRH(D, B - \{a_k\}) \geq NRH(D, B)$, then let $B = B - \{a_k\}$.
 16. End for
 17. Return an optimal reduction set B .
-

3.5. Complexity Analysis of ARNRJE

In the ARNRJE algorithm, suppose that there are m attributes and n samples, and then the calculation of neighborhood classes is frequent in neighborhood decision systems. The process of deriving the neighborhood classes has a great influence on the time complexity of selecting attributes. Notably, the main computation of ARNRJE includes two aspects: obtaining neighborhood classes and computing the neighborhood roughness joint entropy. Then, the buckets sorting algorithm [38] is introduced to further reduce the time complexity of neighborhood classes, and the time complexity of computing neighborhood classes should be $O(mn)$. Meanwhile, the time complexity of calculating neighborhood roughness joint entropy is $O(n)$. Since $O(n) < O(mn)$, the complexity of computing neighborhood roughness joint entropy should be $O(mn)$. Because there exist two loops at steps 3 and 8 of the ARNRJE algorithm, in the worst case, the time complexity of ARNRJE algorithm is $O(m^3n)$. As known that, in the process of achieving attribute reduction task, we usually select a litter of attributes. Suppose that the number of selected attributes is m_R , and in the computation of neighborhood classes, we only need to consider the candidate attributes without touching on the whole attribute set. So, the complexity of computing neighborhood classes is decreased to $O(m_Rn)$. For the ARNRJE algorithm, the times of the outer loop are m and the times of the inner loop are $m - m_R$. Thus, the time complexity of the ARNRJE algorithm is $O(m_Rn(m - m_R)m)$. It is obvious that $m_R \ll m$ in most cases. Therefore, the time complexity of ARNRJE algorithm is approximately $O(mn)$. So far, ARNRJE appears to be more efficient than some of the existing algorithms for attribute reduction in [18,36,45,52,53] for neighborhood decision systems. Furthermore, its space complexity is $O(mn)$.

3.6. An Illustrative Example

In the following, the performance of ARNRJE algorithm is shown through an illustrative example in [20]. A neighborhood decision system $NDS = \langle U, C \cup D, \delta \rangle$ is employed, where $U = \{x_1, x_2, x_3, x_4\}$, $C = \{a, b, c\}$, and $D = \{d\}$ with the values $\{Y, N\}$. The neighborhood decision system is shown in Table 1.

Table 1. A neighborhood decision system.

U	a	b	c	d
x_1	0.12	0.41	0.61	Y
x_2	0.21	0.15	0.14	Y
x_3	0.31	0.11	0.26	N
x_4	0.61	0.13	0.23	N

For Table 1, we use Algorithm 1 for attribute reduction. Given $\delta = 0.3$, the attribute reduction steps are as follows:

- (1) Initialize $B = \emptyset$.
- (2) Let $B = C = \{a, b, c\}$, and the Euclidean distance function is used to calculate the distance between any two objects as follows:

$$\Delta_B(x_1, x_2) = 0.54, \Delta_B(x_1, x_3) = 0.35, \Delta_B(x_1, x_4) = 0.68, \Delta_B(x_2, x_3) = 0.16, \Delta_B(x_2, x_4) = 0.41, \text{ and } \Delta_B(x_3, x_4) = 0.302.$$

Then, we can get the following neighborhood classes:

$$n_B^\delta(x_1) = \{x_1\}, n_B^\delta(x_2) = \{x_2, x_3\}, n_B^\delta(x_3) = \{x_2, x_3\}, \text{ and } n_B^\delta(x_4) = \{x_4\}.$$

Under the equivalence relation, $U/d = \{d_1, d_2\} = \{\{x_1, x_2\}, \{x_3, x_4\}\}$, the Euclidean distance function is Δ_B , and then the upper and lower approximation sets of attribute subset B about d_1 and d_2 are calculated, respectively, by

$$\overline{B}(d_1)_\delta = \{x_1, x_2, x_3\}, \underline{B}(d_1)_\delta = \{x_1\}; \overline{B}(d_2)_\delta = \{x_2, x_3, x_4\}, \text{ and } \underline{B}(d_2)_\delta = \{x_4\}.$$

$$\text{Thus, one has that } \rho_B^\delta(d_1) = \frac{1}{3}, \gamma_B^\delta(d_1) = 1 - \rho_B^\delta(d_1) = \frac{2}{3}, \rho_B^\delta(d_2) = \frac{1}{3}, \text{ and } \gamma_B^\delta(d_2) = 1 - \rho_B^\delta(d_2) = \frac{2}{3}.$$

$$\text{It follows that } NRH(D, C) = -\frac{1}{|U|} \sum_{j=1}^2 \log_2(2^{-\gamma_C^\delta(d_j)}) \times \int_{x_i \in U} \log_2\left(\frac{m(n_C^\delta(x_i) \cap d_j)}{|U|}\right) dx = 0.83.$$

- (3) For all the attributes in C , the attribute significance is calculated by $sig^{inner}(a, C, D) = 0.5081$, $sig^{inner}(b, C, D) = 0$, and $sig^{inner}(c, C, D) = 0.2075$. Since $sig^{inner}(a, C, D) > 0$ and $sig^{inner}(c, C, D) > 0$, then one has $B = \{a, c\}$.
- (4) Since $B = \{a, c\} \neq \emptyset$, by computing, then $NRH(D, B) = 0.83$.
- (5) Because $NRH(D, B) = NRH(D, C)$, it is obtained that the reduction set $B = \{a, c\}$.
- (6) One computes $NRH(D, B - \{a\}) = NRH(D, \{c\}) = 0.3219$ and $NRH(D, B - \{c\}) = NRH(D, \{a\}) = 0.2925$. Since $NRH(D, B - \{a\}) \leq NRH(D, B)$ and $NRH(D, B - \{c\}) \leq NRH(D, B)$, then $B = \{a, c\}$ holds.
- (7) Return the reduction attribute subset $B = \{a, c\}$.

4. Experimental Results and Analysis

4.1. Experiment Preparation

The objective of an attribute reduction method usually includes two aspects: one is to select a small number of attributes and the other is to maintain high classification accuracy. To verify the classification performances of our proposed attribute reduction method described in Subsection 3.4, the comprehensive results of all contrasted algorithms can be achieved and analyzed on nine public data sets (five UCI data sets and four gene expression data sets). The selected five UCI data sets with low-dimensional attributes include the Wine, Sonar, Segmentation, Wdbc, and Wdbc data sets [54]. The selected four gene expression data sets with high-dimensional attributes include the Prostate, DLBCL, Leukemia and Tumors data sets [55]. All the data sets are summarized in Table 2.

Table 2. Description of the seven public data sets.

No.	Data sets	Samples	Attributes	Classes	Author
1	Wine	178	13	3	Faris et al. [56]
2	Sonar	208	60	2	Wang and Li [57]
3	Segmentation	2310	19	7	Liu et al. [38]
4	Wdbc	569	31	2	Li et al. [39]
5	Wpbc	198	33	2	Chen et al. [13]
6	Prostate	136	12600	2	Mu et al. [24]
7	DLBCL	77	5469	2	Sun et al. [8]
8	Leukemia	72	11225	3	Sun et al. [23]
9	Tumors	327	12558	7	Wang et al. [45]

The experiments were performed on a personal computer running Windows 10 with an Intel(R) Core(TM) i5-6500 CPU operating at 3.20 GHz, and 4.0 GB memory. All the simulation experiments were implemented in MATLAB 2016a programming software, and two different classifiers (KNN and LibSVM) were selected to verify the classification accuracy in WEKA software, where the parameter k in KNN was set to 3 and the linear kernel functions were selected in LibSVM. All of the following experimental comparisons for classification on the selected attributes are implemented using a 10-fold cross-validation with all the test data sets, where every data set is first randomly divided into ten portions which are the same size subset of data each other, one data subset is used as the testing data set, the rest nine data subsets are used as the training data set, and each of the ten data subsets only is employed exactly once as the testing data set; secondly, the operation of the cross-validation is repeated ten times; finally, the average of ten test results is as the obtained classification accuracy [45].

4.2. Effect of Different Neighborhood Parameter Values

The following part of our experiments concerns the reduction rate and the classification accuracy under the different neighborhood parameter values. The reduction rate and the classification accuracy of an attribute subset for the different neighborhood parameter values are discussed to obtain a suitable

neighborhood parameter value and a subset of attributes. Chen et al. [20] adopted a reduction rate for evaluating the attribute redundancy degree of attribute reduction algorithms, which is described as:

$$Rate_{\delta} = \frac{|C| - |R|}{|C|}, \tag{27}$$

where $|C|$ is the number of all of the conditional attributes in the data set, and $|R|$ is the number of the reduced attributes obtained under the different neighborhood parameter values.

For the four high-dimensional gene expression data sets (Prostate, DLBCL, Leukemia and Tumors), the Fisher score method [8] is used for preliminary dimension reduction. For each gene expression data set, the Fisher score method is used to calculate the Fisher score value and sequence it for each gene, and l genes are selected to construct a candidate gene subset. The classification accuracy under different dimensions is obtained by using WEKA software, so that the appropriate dimension can be selected for the attribute reduction algorithm. Figure 2 shows the changing trend of the classification accuracy versus the number of genes on the four gene expression data sets.

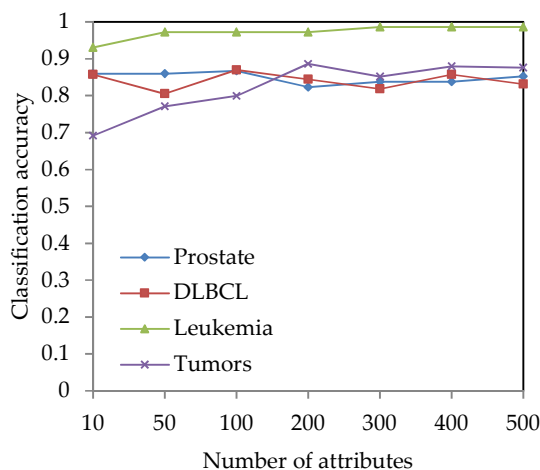


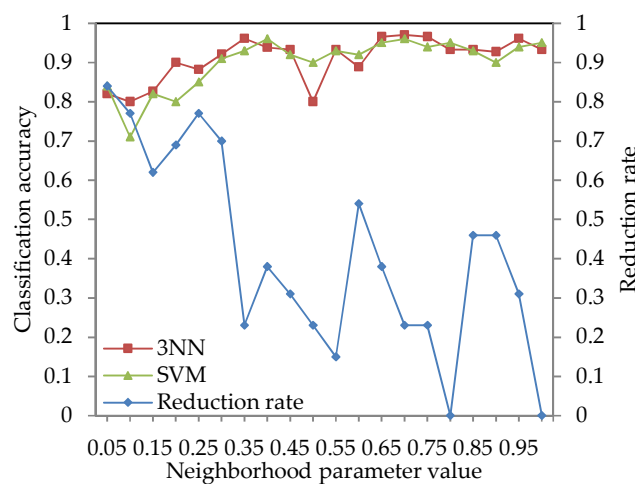
Figure 2. The classification accuracy versus the number of genes on the four gene expression data sets.

From Figure 2, it can be seen that when the number of genes increases, the classification accuracy is also changed. Since the cardinality of selected genes and the classification accuracy for selected genes are all important, they are two indices for evaluating the classification performance of attribute reduction methods. Then, the appropriate values of l need to be selected from Figure 2. Hence, the values of l are set to 50-dimension and 100-dimension for the Prostate and DLBCL data set, respectively. For the Leukemia and Tumors data sets, the values of l can be set to 300-dimension and 200-dimension, respectively.

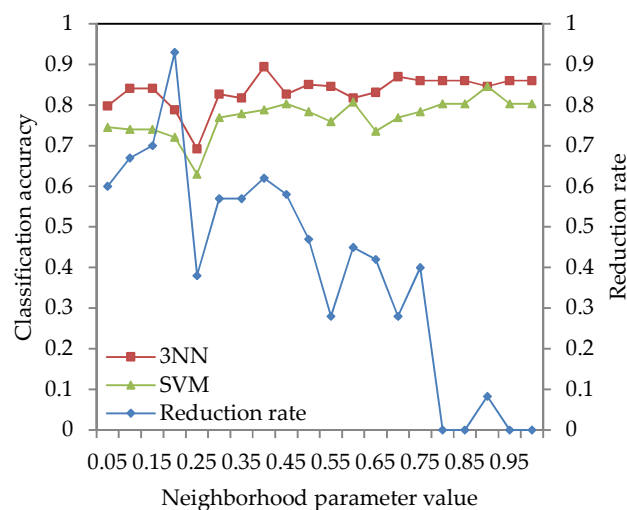
The classification accuracy of data sets given in Table 2 was obtained by using the ARNRJE algorithm with different neighborhood parameter values. After obtaining the results of attribute reduction with different parameters, WEKA is used to obtain the classification accuracy under the 3NN and LibSVM classifiers. The results are shown in Figure 3, where the horizontal coordinate denotes the neighborhood parameters with $\delta \in [0.05, 1]$ at intervals of 0.05, and the left and the right vertical coordinate represent the classification accuracy and the reduction rate, respectively.

Figure 3 shows that the classification accuracy of selected attributes by the ARNRJE algorithm is increasing, and the reduction rate is decreasing with the neighborhood parameter values changing from 0.05 to 1 in most cases. It is easily observed that the different neighborhood parameter values have great influence on the classification performance of ARNRJE. Then, this illustrates that the thinner the granule is, the smaller the roughness of the granule is when the values of different neighborhood parameter are smaller. It follows that the reduction rate increases as the roughness of the granule decreases. Thus, the optimal neighborhood parameter values can be selected for each data set. Figure 3a

shows the classification accuracy of the Wine data set with different neighborhood parameter values, and the neighborhood parameter can be set to 0.65. Figure 3b demonstrates the classification accuracy of the Sonar data set with different neighborhood parameter values. The reduction rate decreases as the neighborhood parameter values increase, and the classification accuracy of selected attributes reaches the relative maximum value when the neighborhood parameter is 0.4. So, the neighborhood parameter of the Sonar data set can be set to 0.4. From Figure 3c, it can be seen that the classification accuracy reaches the relative best performance when the parameter equals 0.75 on the Segmentation data set. Figure 3d displays the classification accuracy of the Wdbc data set for the different neighborhood parameters. The neighborhood parameter of the Wdbc data set can be set to 0.35. Similar to Figure 3b, Figure 3e reveals the classification accuracy of selected attributes reaching a relative maximum when the parameter is set as 0.55 for Wpbc data set. For Prostate and DLBCL with the different neighborhood parameter values in Figure 3f,g, the neighborhood parameter of the Prostate data set can be set to 0.9, and that of the DLBCL data set should be set to 0.5. Figure 3h,i demonstrates that the reduction rate is decreasing as the parameters increase in most situations, and the neighborhood parameter of Leukemia and Tumors data sets can be set to 0.4 and 0.25, respectively.

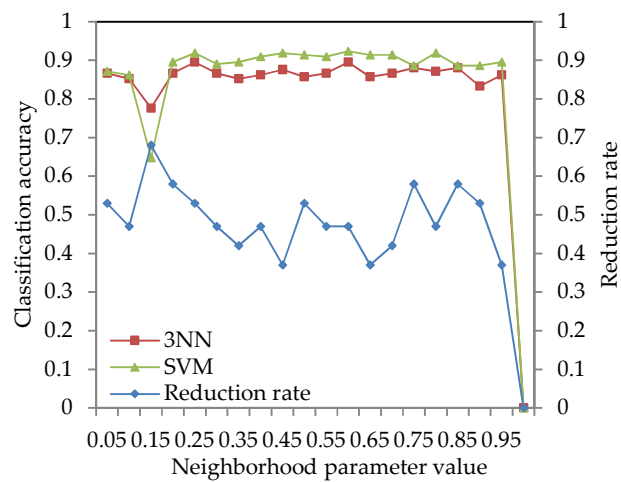


(a) Wine data set

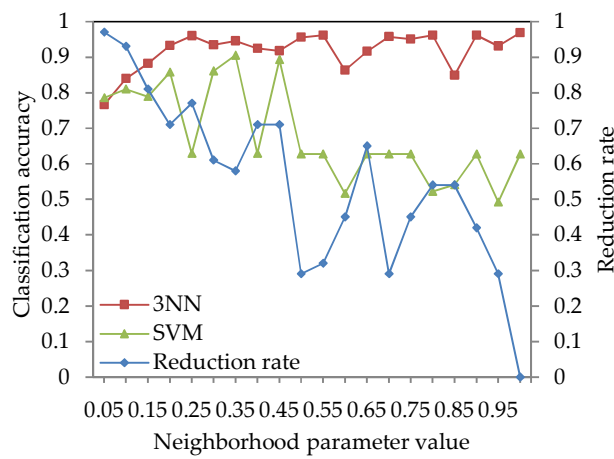


(b) Sonar data set

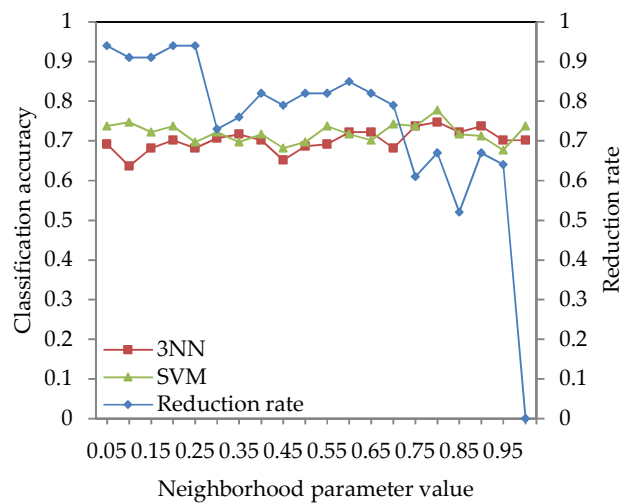
Figure 3. Cont.



(c) Segmentation data set

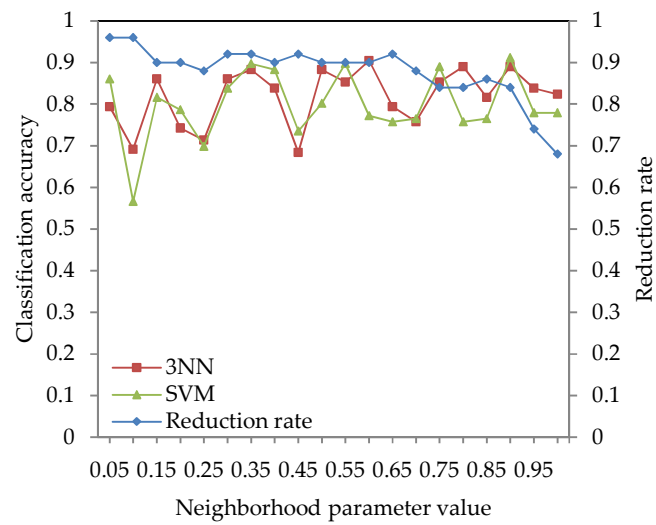


(d) Wdbc data set

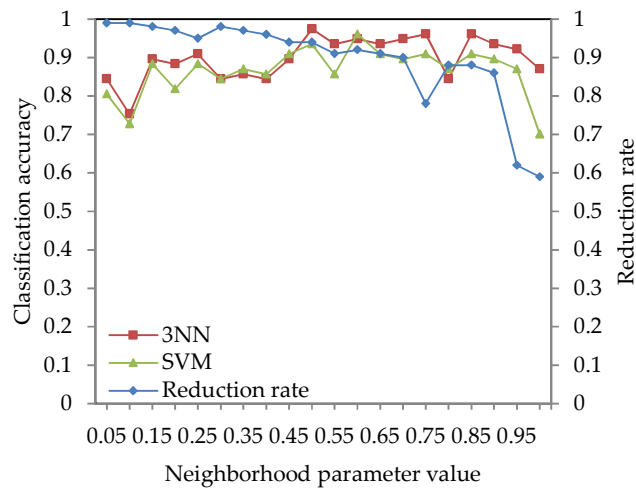


(e) Wpbc data set

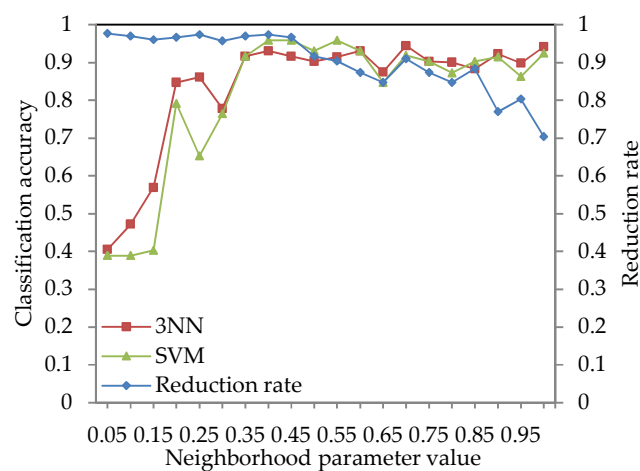
Figure 3. Cont.



(f) Prostate data set



(g) DLBCL data set



(h) Leukemia data set

Figure 3. Cont.

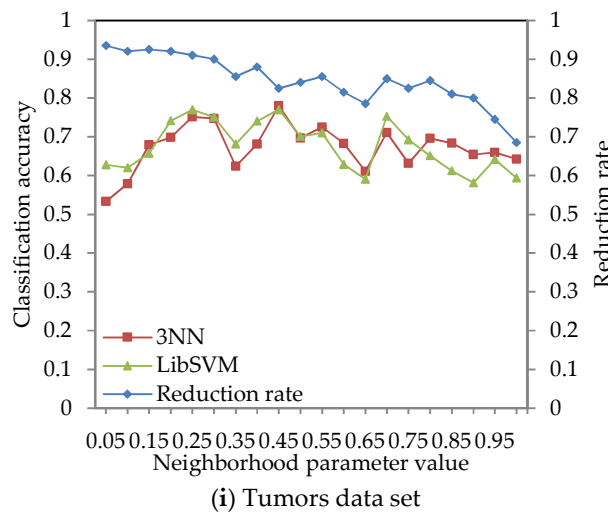


Figure 3. Reduction rate and classification accuracy for nine data sets with neighborhood parameter values.

4.3. Comparison of Reduction Results with Three Related Reduction Algorithms

This portion of our experiments evaluates the performance of our proposed ARNRJE algorithm in terms of the selected attribute subset of data sets. The ARNRJE algorithm is compared with the following two reduction algorithms: (1) the fuzzy information entropy-based feature selection algorithm (FINEN) [52,58], and (2) the neighborhood entropy-based feature selection algorithm (NEIEN) [18]. By using the neighborhood parameters where the classification accuracy is obtained in Subsection 4.2, the attribute reduction results and the number of selected attributes on the nine data sets from Table 2 are shown in Table 3.

Table 3. The reduction results and the number of selected attributes with the three reduction algorithms.

Data Sets	FINEN	NEIEN	ARNRJE	δ
Wine	{1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13}/12	{1, 2, 3, 4, 5, 7, 10, 11, 12, 13}/10	{1, 2, 3, 5, 7, 10, 11, 13}/8	0.65
Sonar	{1, 5, 9, 10, 11, 12, 18, 19, 22, 26, 27, 28, 29, 32, 35, 36, 37, 40, 45, 46, 48, 53, 57, 58, 59, 60}/26	{6, 10, 11, 12, 15, 17, 18, 20, 21, 23, 24, 26, 28, 29, 30, 32, 33, 36, 37, 39, 40, 41, 42, 45, 48, 50, 54, 57}/29	{2, 3, 4, 5, 9, 10, 11, 12, 13, 14, 16, 22, 24, 30, 32, 34, 36, 37, 38, 39, 46, 57, 60}/23	0.4
Segmentation	{2, 5, 6, 7, 11, 12, 13, 17, 18}/9	{2, 5, 6, 7, 11, 12, 13, 17, 18}/9	{2, 5, 6, 10, 11, 13, 17, 18}/8	0.75
Wdbc	{7, 8, 10, 12, 13, 16, 21, 22, 25, 27, 28, 29}/12	{1, 7, 8, 10, 13, 16, 21, 22, 25, 27, 28, 29}/12	{6, 8, 9, 11, 12, 14, 16, 19, 20, 25, 27, 28, 29}/13	0.35
Wpbc	{1, 12, 13, 16, 24, 32}/6	{1, 5, 12, 24, 32}/5	{2, 19, 23, 24, 29, 31}/6	0.55
Prostate	{4483, 6185, 8129, 8623, 8850, 9850, 10753, 12067}/8	{4483, 4847, 6185, 6627, 8623, 8850, 9587, 12067}/8	{11052, 6185, 8986, 5486, 6392, 5757, 8850, 4483}/8	0.9
DLBCL	{453, 1570, 1698, 3127, 3257, 4767}/6	{453, 2930, 3574, 4767, 5283}/5	{453, 1479, 1570, 3127, 3257, 4767}/6	0.5
Leukemia	{2833, 6720, 5555, 10127, 10038, 4839, 8952, 9053}/8	{2833, 6720, 5555, 10127, 10038, 3479, 8964, 515}/8	{461, 1787, 1834, 1962, 2131, 2356, 3821, 5552}/8	0.4
Tumors	{2543, 7648, 3264, 6320, 5411, 6671, 8548, 7781, 10126, 6764, 4178, 4448, 8337, 3043, 4831, 3880}/16	{5411, 6320, 7648, 3264, 3324, 6671, 4300, 6079, 6764, 10126, 8397, 8383, 9046, 7922, 10865, 8687, 2132}/17	{3880, 843, 1730, 3342, 6151, 2960, 3264, 3596, 5624, 4026, 7648, 8383, 8332, 9788, 5412, 8556, 3324, 10126}/18	0.25

Table 3 lists the selected attribute subsets. It can be seen that the best attribute subsets selected by the FINEN, NEIEN, and ARNRJE algorithms are the same as each other in some situations, and the number of attributes selected by ARNRJE is less than those of FINEN and NEIEN in the majority of cases. For the Wine, Sonar, and Segmentation data sets, the numbers of attributes selected by the three reduction algorithms are different, where the ARNRJE exhibits the best performance. The slight differences of the three data sets may be caused by the fact that the selected attribute subsets are obtained through reducing the whole data set. For the Wdbc, Wpbc, Prostate, DLBCL, Leukemia and Tumors data sets, the selected attribute subsets are different in general, but the numbers of attributes

selected by the three algorithms are very close to each other. Therefore, the proposed ARNRJE algorithm is efficient in dimension reduction of low-dimensional and high-dimensional data sets.

4.4. Comparison of Classification Results with Six Reduction Methods on Two Different Classifiers

To further demonstrate the classification performance of our proposed method, six methods are used to evaluate the classification accuracy on the selected attributes. The ARNRJE algorithm is compared with the five related reduction methods, which include: (1) the original data processing method (ODP), (2) the neighborhood rough set algorithm (NRS) [22], (3) the fuzzy rough dependency constructed by intersection operations of fuzzy similarity relations algorithm (FRSINT) [53], (4) the FINEN algorithm [52,58], and (5) the NEIEN algorithm [18]. The two classifiers (3NN and LibSVM) in WEKA are employed to test the classification accuracy. Table 4 denotes the average sizes of attribute subsets selected by the six methods using 10-fold cross validation. What's more, the corresponding classification accuracy of selected attributes under the 3NN and LibSVM classifiers with 10-fold cross validation is shown in Tables 5 and 6, respectively.

Table 4. Average sizes of attribute subsets selected by the six methods using 10-fold cross validation.

Data Sets	ODP	NRS	FRSINT	FINEN	NEIEN	ARNRJE
Wine	13	9.1	8.1	12.3	10.2	9.5
Sonar	60	24.8	18.7	25.8	28.9	24.6
Segmentation	19	10.7	8.4	9.5	9.2	8.9
Wdbc	30	17.3	11.9	12.1	11.8	13.3
Wdbc	32	11.6	7.8	6.4	5.3	6.2
Prostate	12600	6.5	8.9	8.4	7.7	8
DLBCL	5469	8.3	8.8	6.1	5.3	7.6
Leukemia	11225	14.7	9.8	8.5	8.2	9.1
Tumors	12558	10.6	9.5	15.8	17.1	18.2
Average	4667.3	12.6	10.2	11.7	11.5	11.7

From Table 4, comparing the average sizes of selected attribute subsets by using 10-fold cross validation, the FRSINT, NEIEN, and ARNRJE algorithms are obviously superior to the RS and FINEN algorithms, but the ARNRJE algorithm is slightly inferior to the FRSINT and NEIEN algorithms. From Tables 5 and 6, the difference among the six methods can be clearly identified. Then, it can be clearly observed that the classification accuracy of the proposed ARNRJE algorithm outperforms that of the other five methods on most of the nine data sets, except for the Segmentation, Wdbc, and Tumors data sets under the 3NN classifier and the Sonar, Segmentation, Wdbc, and Tumors data sets under the LibSVM classifier. Furthermore, the average classification accuracy of the ARNRJE is the highest on the LibSVM classifier, and has greatly improvement, but the ARNRJE is 0.2% lower than that of FINEN in classification accuracy on the 3NN classifier. From Tables 4 and 5 under 3NN classifier, although ARNRJE is not as well as FRSINT and NEIEN in the average sizes of selected attribute subsets, the classification accuracy of ARNRJE is nearly 2%–6% higher than that of FRSINT, and that of ARNRJE is approximately 1%–6% higher than that of NEIEN, except for the Segmentation data set. In addition, the classification performance of ARNRJE is better than that of NRS and FINEN on the whole. Though there is some difference in the number of attributes selected by ARNRJE, NRS and FINEN, the accuracy of ARNRJE is higher than that of NRS and FINEN, except for the Wdbc, Segmentation, and Tumors data sets, respectively. The reason is that some important information attributes of the Wdbc and Segmentation data sets are lost in the process of reduction for ARNRJE. Similarly, as seen from Tables 4 and 6, under the LibSVM classifier, the classification accuracy of ARNRJE is 1%–6% higher than that of FRSINT, ARNRJE is 2%–6% higher than that of NEIEN in classification accuracy, and compared with FINEN, the accuracy of ARNRJE is 0.5%–5% higher, except for the Sonar, Segmentation, Wdbc, and Tumors data sets. For the Wdbc data set, the classification accuracy of ARNRJE is 4% lower than NRS, but ARNRJE selects the less attributes than NRS, and exhibits the better classification accuracy

than that obtained by NRS on the other six data sets. As far as the average classification accuracy is concerned, our ARNRJE algorithm shows great stability on the 3NN and LibSVM classifiers, whereas the classification accuracy of the ODP, NRS, and FINEN algorithms is slightly unstable. Based on the results in Table 4, it can be seen that for the Sonar, Segmentation, Wdbc, and Tumors data sets, the proposed ARNRJE method reduces some important attributes in the process of reduction, resulting in the decrease of classification accuracy of reduction sets with fewer attributes. The above results show that no algorithm is congruously better than the others for different learning tasks and classifiers. Overall speaking, our proposed approach can obviously reduce the redundant data and outperforms the other related attribute reduction methods. The experimental results show that our method is an efficient reduction method for redundant data sets, and can improve the classification accuracy for most of the data sets.

Table 5. Classification accuracy of the six methods under the 3NN classifier.

Data Sets	ODP	NRS	FRSINT	FINEN	NEIEN	ARNRJE
Wine	0.9192	0.9453	0.9281	0.9577	0.9620	0.9775
Sonar	0.8605	0.8588	0.8504	0.8513	0.8326	0.8942
Segmentation	0.8714	0.8021	0.9506	0.9504	0.9488	0.8381
Wdbc	0.9432	0.9553	0.9366	0.9226	0.9385	0.9456
Wpbc	0.6667	0.6752	0.6312	0.6613	0.6263	0.6919
Prostate	0.8235	0.8329	0.8503	0.8689	0.8431	0.8897
DLBCL	0.8831	0.9610	0.9635	0.9635	0.9585	0.9740
Leukemia	0.7339	0.9274	0.8655	0.9246	0.886	0.9306
Tumors	0.7074	0.725	0.7239	0.7781	0.7372	0.7194
Average	0.8232	0.8537	0.8556	0.8754	0.8592	0.8734

Table 6. Classification accuracy of the six methods under the LibSVM classifier.

Data Sets	ODP	NRS	FRSINT	FINEN	NEIEN	ARNRJE
Wine	0.9210	0.9213	0.9295	0.9503	0.9210	0.9719
Sonar	0.6587	0.7735	0.7909	0.8168	0.8036	0.7885
Segmentation	0.9048	0.8606	0.9438	0.9317	0.9356	0.9095
Wdbc	0.5167	0.9453	0.9362	0.9230	0.9449	0.9051
Wpbc	0.7374	0.7029	0.7002	0.6875	0.7188	0.7374
Prostate	0.8750	0.8353	0.8527	0.9039	0.8691	0.9118
DLBCL	0.8701	0.9231	0.924	0.9051	0.8758	0.9351
Leukemia	0.4679	0.9165	0.9454	0.9122	0.942	0.9583
Tumors	0.2788	0.7516	0.7308	0.7742	0.7712	0.7627
Average	0.6923	0.8478	0.8615	0.8672	0.8647	0.8756

In the above experiments, the coarse ordering of the five methods on time complexity is as follows: $O(\text{FINEN}) = O(\text{FRSINT}) > O(\text{NRS}) > O(\text{NEIEN}) > O(\text{ARNRJE})$, where $O(A)$ represents the time complexity of A algorithm. The time complexity of NEIEN algorithm is $O(n^2)$ [18]. For the UCI data set with low-dimension, the NEIEN algorithm has the lower time complexity. As we know, the number of samples is usually much greater than that of attributes on the UCI data sets in most cases, whereas for the gene expression data sets, the number of genes is much larger than that of samples. Since the time complexity of ARNRJE algorithm is $O(mn)$, it is less than that of the NEIEN for large-scale and high-dimensional data sets. Although the time complexity of NEIEN is lower than that of ARNRJE on UCI data sets, the classification performance of ARNRJE algorithm is higher than that of NEIEN algorithm in most instances. For the NRS algorithm, the time complexity is $O(m^2 n \log n)$ [22]. Since the time of the FRSINT algorithm is mainly spent on getting the fuzzy-rough membership of each sample for different decision classes, the FRSINT algorithm runs slowly and its time complexity is $O(m^2 n^2)$ [45,53]. In addition, the time complexity of FINEN is also $O(m^2 n^2)$ [52], which is time-consuming. The reason is that since the FINEN algorithm is based on similarity relation, they need a lot of time to calculate the similarity relation of attributes [38]. Therefore, it can be easily

proven that the ARNRJE algorithm has lower time complexity, can effectively reduce the redundancy, improve the classification accuracy, and optimize the classification process of large-scale complex data.

4.5. Comparison of Recall Rate with Three Reduction Methods on Two Different Classifiers

The final portion of our experiments is to measure the recall classification index to evaluate the classification performances of three reduction methods on two different classifiers. The recall rate [59] as a metric is employed to assess the classification performance, which is described as:

$$R = \frac{TP}{TP + FN}, \quad (28)$$

where True Positive (*TP*) denotes the number of positive instances diagnosed correctly, and False Negative (*FN*) represents the number of positive instances detected as negative.

Tables 7 and 8 demonstrate the testing results of the recall rate with the FINEN, NEIEN with ARNRJE on the nine data sets on the 3NN and LibSVM classifiers, respectively. From Tables 7 and 8, the ARNRJE algorithm achieves the highest average recall rate under the two classifiers, and outperforms the FINEN and NEIEN algorithms on most of all the nine data sets. It can be seen from Table 7 under the 3NN classifier that ARNRJE is nearly 4% lower than FINEN for Prostate, and 3% for Tumors in recall rate, respectively. From Table 8 under the LibSVM classifier, ARNRJE is slightly inferior to FINEN for Wdbc and NEIEN for Tumors. The reason is that some important information attributes are lost in the process of preliminary dimensionality reduction or reduction for ARNRJE for the Prostate, DLBCL, Wdbc and Tumors data sets. Thus, this causes misclassification of conditional attributes, and leads to the slightly lower recall rate. The above results manifest that for different learning tasks and classifiers, no algorithm can consistently superior to other algorithms. In general, our proposed ARNRJE algorithm has a relatively good classification performance by measuring the recall rate.

Table 7. The recall rate with the three reduction algorithms under 3NN.

Data Sets	FINEN	NEIEN	ARNRJE
Wine	0.961	0.978	1.000
Sonar	0.910	0.937	0.955
Segmentation	0.833	0.833	0.838
Wdbc	0.951	0.953	0.966
Wdbc	0.815	0.753	0.849
Prostate	0.934	0.882	0.890
DLBCL	1.000	0.974	0.974
Leukemia	0.917	0.958	0.979
Tumors	0.567	0.733	0.867
Average	0.876	0.889	0.924

Table 8. The recall rate with the three reduction algorithms under LibSVM.

Data Sets	FINEN	NEIEN	ARNRJE
Wine	0.958	1.000	1.000
Sonar	0.771	0.688	0.874
Segmentation	0.567	0.567	0.667
Wdbc	0.930	0.627	0.924
Wdbc	0.737	0.737	0.768
Prostate	0.566	0.566	0.566
DLBCL	0.753	0.753	1.000
Leukemia	0.389	0.389	0.653
Tumors	0.567	0.691	0.566
Average	0.693	0.669	0.780

5. Conclusions

Attribute reduction, one of the important steps in classification learning, can improve the classification performance in most of cases and decrease the cost of classification. Uncertainty measures for calculating distinguishing ability of attribute subsets play an important role in the process of attribute reduction. The neighborhood rough model can effectively solve the reduction problem of numerical and continuous-valued information system. In this paper, a neighborhood rough sets-based attribute reduction method using Lebesgue and entropy measures is proposed to improve the classification performance of continuous data set. Based on Lebesgue and entropy measures, some neighborhood entropy-based uncertainty measures in neighborhood decision systems is investigated. Then, the neighborhood roughness joint entropy is presented for handling the uncertainty and noisy of neighborhood decision systems, which combines the algebraic view and the information view in neighborhood rough sets. Moreover, their corresponding properties and relationships are discussed. Thus, a heuristic search algorithm is constructed to improve the computational efficiency of selected attributes in neighborhood decision systems. The experimental results show that our proposed algorithm can obtain a small, effective attribute subset with great classification performance.

Author Contributions: L.S. and J.X. conceived the methodology and designed the experiments; L.W. achieved the algorithms and implemented the experiments; S.Z. analyzed the results; and L.W. drafted the manuscript. All authors read and revised the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (Grants 61772176, 61402153 and 61370169), the China Postdoctoral Science Foundation (Grant 2016M602247), the Plan for Scientific Innovation Talent of Henan Province (Grant 184100510003), the Key Scientific and Technological Project of Henan Province (Grant 182102210362), the Young Scholar Program of Henan Province (Grant 2017GGJS041), the Natural Science Foundation of Henan Province (Grants 182300410130, 182300410368 and 182300410306), and the Ph.D. Research Foundation of Henan Normal University (Grant qd15132).

Conflicts of Interest: The authors declare no conflict of interests.

References

1. Wang, Q.; Qian, Y.H.; Liang, X.Y.; Guo, Q.; Liang, J.Y. Local neighborhood rough set. *Knowl.-Based Syst.* **2018**, *135*, 53–64. [[CrossRef](#)]
2. Gao, C.; Lai, Z.H.; Zhou, J.; Zhao, C.R.; Miao, D.Q. Maximum decision entropy-based attribute reduction in decision-theoretic rough set model. *Knowl.-Based Syst.* **2018**, *143*, 179–191. [[CrossRef](#)]
3. Zhang, X.Y.; Miao, D.Q. Quantitative/qualitative region-change uncertainty/certainty in attribute reduction: Comparative region-change analyses based on granular computing. *Inf. Sci.* **2016**, *334*, 174–204. [[CrossRef](#)]
4. Dong, H.B.; Li, T.; Ding, R.; Sun, J. A novel hybrid genetic algorithm with granular information for feature selection and optimization. *Appl. Soft Comput.* **2018**, *65*, 33–46. [[CrossRef](#)]
5. Hu, L.; Gao, W.F.; Zhao, K.; Zhang, P.; Wang, F. Feature selection considering two types of feature relevancy and feature interdependency. *Expert Syst. Appl.* **2018**, *93*, 423–434. [[CrossRef](#)]
6. Liu, Y.; Xie, H.; Tan, K.Z.; Chen, Y.H.; Xu, Z.; Wang, L.G. Hyperspectral band selection based on consistency-measure of neighborhood rough set theory. *Meas. Sci. Technol.* **2016**, *27*, 055501. [[CrossRef](#)]
7. Lyu, H.Q.; Wan, M.X.; Han, J.Q.; Liu, R.L.; Wang, C. A filter feature selection method based on the Maximal Information Coefficient and Gram-Schmidt Orthogonalization for biomedical data mining. *Comput. Biol. Med.* **2017**, *89*, 264–274. [[CrossRef](#)] [[PubMed](#)]
8. Sun, L.; Zhang, X.Y.; Qian, Y.H.; Xu, J.C.; Zhang, S.G.; Tian, Y. Joint neighborhood entropy-based gene selection method with fisher score for tumor classification. *Appl. Intell.* **2018**. [[CrossRef](#)]
9. Jadhav, S.; He, H.M.; Jenkins, K. Information gain directed genetic algorithm wrapper feature selection for credit rating. *Appl. Soft Comput.* **2018**, *69*, 541–553. [[CrossRef](#)]
10. Mariello, A.; Battiti, R. Feature selection based on the neighborhood entropy. *IEEE Trans. Neural Netw. Learn. Syst.* **2018**, *99*, 1–10. [[CrossRef](#)]
11. Das, A.K.; Sengupta, S.; Bhattacharyya, S. A group incremental feature selection for classification using rough set theory based genetic algorithm. *Appl. Soft Comput.* **2018**, *65*, 400–411. [[CrossRef](#)]

12. Imani, M.B.; Keyvanpour, M.R.; Azmi, R. A novel embedded feature selection method: A comparative study in the application of text categorization. *Appl. Artif. Intell.* **2013**, *27*, 408–427. [[CrossRef](#)]
13. Chen, Y.M.; Zeng, Z.Q.; Lu, J.W. Neighborhood rough set reduction with fish swarm algorithm. *Soft Comput.* **2017**, *21*, 6907–6918. [[CrossRef](#)]
14. Li, B.Y.; Xiao, J.M.; Wang, X.H. Feature reduction for power system transient stability assessment based on neighborhood rough set and discernibility matrix. *Energies* **2018**, *11*, 185. [[CrossRef](#)]
15. Gu, X.P.; Li, Y.; Jia, J.H. Feature selection for transient stability assessment based on kernelized fuzzy rough sets and memetic algorithm. *Int. J. Electr. Power Energy Syst.* **2015**, *64*, 664–670. [[CrossRef](#)]
16. Raza, M.S.; Qamar, U. A parallel rough set based dependency calculation method for efficient feature selection. *Appl. Soft Comput.* **2018**, *71*, 1020–1034. [[CrossRef](#)]
17. Luan, X.Y.; Li, Z.P.; Liu, T.Z. A novel attribute reduction algorithm based on rough set and improved artificial fish swarm algorithm. *Neurocomputing* **2015**, *174*, 522–529. [[CrossRef](#)]
18. Hu, Q.H.; Zhang, L.; Zhang, D.; Pan, W.; An, S.; Pedrycz, W. Measuring relevance between discrete and continuous features based on neighborhood mutual information. *Expert Syst. Appl.* **2011**, *38*, 10737–10750. [[CrossRef](#)]
19. Chakraborty, D.B.; Pal, S.K. Neighborhood rough filter and intuitionistic entropy in unsupervised tracking. *IEEE Trans. Fuzzy Syst.* **2018**, *26*, 2188–2200. [[CrossRef](#)]
20. Chen, Y.M.; Zhang, Z.J.; Zheng, J.Z.; Ma, Y.; Xue, Y. Gene selection for tumor classification using neighborhood rough sets and entropy measures. *J. Biomed. Inf.* **2017**, *67*, 59–68. [[CrossRef](#)]
21. Hu, Q.H.; Pan, W.; An, S.; Ma, P.J.; Wei, J.M. An efficient gene selection technique for cancer recognition based on neighborhood mutual information. *Int. J. Mach. Learn. Cybern.* **2010**, *1*, 63–74. [[CrossRef](#)]
22. Hu, Q.H.; Yu, D.; Liu, J.F.; Wu, C.X. Neighborhood rough set based heterogeneous feature subset selection. *Inf. Sci.* **2008**, *178*, 3577–3594. [[CrossRef](#)]
23. Sun, L.; Zhang, X.Y.; Xu, J.C.; Wang, W.; Liu, R.N. A gene selection approach based on the Fisher linear discriminant and the neighborhood rough set. *Bioengineered* **2018**, *9*, 144–151. [[CrossRef](#)] [[PubMed](#)]
24. Mu, H.Y.; Xu, J.C.; Wang, Y.; Sun, L. Feature genes selection using Fisher transformation method. *J. Intell. Fuzzy Syst.* **2018**, *34*, 4291–4300. [[CrossRef](#)]
25. Halmos, P.R. *Measure Theory*; Litton Educational Publishing, Inc. and Springer-Verlag New York Inc.: New York, NY, USA, 1970; pp. 100–152.
26. Song, J.J.; Li, J. Lebesgue theorems in non-additive measure theory. *Fuzzy Sets Syst.* **2005**, *149*, 543–548. [[CrossRef](#)]
27. Xu, X.; Lu, Z.Z.; Luo, X.P. A kernel estimate method for characteristic function-based uncertainty importance measure. *Appl. Math. Model.* **2017**, *42*, 58–70. [[CrossRef](#)]
28. Halčinová, L.; Hutník, O.; Kiseľák, J.; Šupina, J. Beyond the scope of super level measures. *Fuzzy Sets Syst.* **2018**. [[CrossRef](#)]
29. Park, S.R.; Kolouri, S.; Kundu, S.; Rohde, G.K. The cumulative distribution transform and linear pattern classification. *Appl. Comput. Harmon. Anal.* **2018**, *45*, 616–641. [[CrossRef](#)]
30. Marzio, M.D.; Fensore, S.; Panzera, A.; Taylor, C.C. Local binary regression with spherical predictors. *Stat. Probab. Lett.* **2019**, *144*, 30–36. [[CrossRef](#)]
31. Khanjani Shiraz, R.K.; Fukuyama, H.; Tavana, M.; Di Caprio, D. An integrated data envelopment analysis and free disposal hull framework for cost-efficiency measurement using rough sets. *Appl. Soft Comput.* **2016**, *46*, 204–219. [[CrossRef](#)]
32. Zhang, Z.F.; David, J. An entropy-based approach for assessing the operation of production logistics. *Expert Syst. Appl.* **2019**, *119*, 118–127. [[CrossRef](#)]
33. Wang, C.Z.; He, Q.; Shao, M.W.; Xu, Y.Y.; Hu, Q.H. A unified information measure for general binary relations. *Knowl.-Based Syst.* **2017**, *135*, 18–28. [[CrossRef](#)]
34. Ge, H.; Yang, C.J.; Li, L.S. Positive region reduct based on relative discernibility and acceleration strategy. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.* **2018**, *26*, 521–551. [[CrossRef](#)]
35. Sun, L.; Xu, J.C. A granular computing approach to gene selection. *Bio-Med. Mater. Eng.* **2014**, *24*, 1307–1314.
36. Fan, X.D.; Zhao, W.D.; Wang, C.Z.; Huang, Y. Attribute reduction based on max-decision neighborhood rough set model. *Knowl.-Based Syst.* **2018**, *151*, 16–23. [[CrossRef](#)]
37. Meng, J.; Zhang, J.; Li, R.; Luan, Y.S. Gene selection using rough set based on neighborhood for the analysis of plant stress response. *Appl. Soft Comput.* **2014**, *25*, 51–63. [[CrossRef](#)]

38. Liu, Y.; Huang, W.L.; Jiang, Y.L.; Zeng, Z.Y. Quick attribute reduct algorithm for neighborhood rough set model. *Inf. Sci.* **2014**, *271*, 65–81. [[CrossRef](#)]
39. Li, W.W.; Huang, Z.Q.; Jia, X.Y.; Cai, X.Y. Neighborhood based decision-theoretic rough set models. *Int. J. Approx. Reason.* **2016**, *69*, 1–17. [[CrossRef](#)]
40. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Techn. J.* **1948**, *27*, 379–423. [[CrossRef](#)]
41. Chen, Y.M.; Xue, Y.; Ma, Y.; Xu, F.F. Measures of uncertainty for neighborhood rough sets. *Knowl.-Based Syst.* **2017**, *120*, 226–235. [[CrossRef](#)]
42. Liang, J.Y.; Shi, Z.Z.; Li, D.Y.; Wierman, M.J. Information entropy, rough entropy and knowledge granulation in incomplete information systems. *Int. J. Gen. Syst.* **2006**, *35*, 641–654. [[CrossRef](#)]
43. Sun, L.; Xu, J.C.; Tian, Y. Feature selection using rough entropy-based uncertainty measures in incomplete decision systems. *Knowl.-Based Syst.* **2012**, *36*, 206–216. [[CrossRef](#)]
44. Liu, Y.; Xie, H.; Chen, Y.H.; Tan, K.Z.; Wu, X. Neighborhood mutual information and its application on hyperspectral band selection for classification. *Chemom. Intell. Lab. Syst.* **2016**, *157*, 140–151. [[CrossRef](#)]
45. Wang, C.Z.; Hu, Q.H.; Wang, X.Z.; Chen, D.G.; Qian, Y.H.; Dong, Z. Feature selection based on neighborhood discrimination index. *IEEE Trans. Neural Netw. Learn. Syst.* **2018**, *29*, 2986–2999. [[CrossRef](#)]
46. Wang, G.Y. Rough reduction in algebra view and information view. *Int. J. Intell. Syst.* **2003**, *18*, 679–688. [[CrossRef](#)]
47. Teng, S.H.; Lu, M.; Yang, A.F.; Zhang, J.; Nian, Y.J.; He, M. Efficient attribute reduction from the viewpoint of discernibility. *Inf. Sci.* **2016**, *326*, 297–314. [[CrossRef](#)]
48. Qian, Y.H.; Liang, J.Y.; Pedryce, W.; Dang, C.Y. Positive approximation: An accelerator for attribute reduction in rough set theory. *Artif. Intell.* **2010**, *174*, 597–618. [[CrossRef](#)]
49. Pawlak, Z.; Skowron, A. Rudiments of rough sets. *Inf. Sci.* **2007**, *177*, 3–27. [[CrossRef](#)]
50. Sun, L.; Xu, J.C. Information entropy and mutual information-based uncertainty measures in rough set theory. *Appl. Math. Inf. Sci.* **2014**, *8*, 1973–1985. [[CrossRef](#)]
51. Chen, Y.M.; Wu, K.S.; Chen, X.H.; Tang, C.H.; Zhu, Q.X. An entropy-based uncertainty measurement approach in neighborhood systems. *Inf. Sci.* **2014**, *279*, 239–250. [[CrossRef](#)]
52. Hu, Q.H.; Yu, D.; Xie, Z.X.; Liu, J.F. Fuzzy probabilistic approximation spaces and their information measures. *IEEE Trans. Fuzzy Syst.* **2006**, *14*, 191–201.
53. Jensen, R.; Shen, Q. Semantics-preserving dimensionality reduction: Rough and fuzzy-rough-based approaches. *IEEE Trans. Knowl. Data Eng.* **2004**, *16*, 1457–1471. [[CrossRef](#)]
54. UCI Machine Learning Repository. Available online: <http://archive.ics.uci.edu/ml/index.php> (accessed on 15 December 2018).
55. BROAD INSTITUTE, Cancer Program Legacy Publication Resources. Available online: <http://portals.broadinstitute.org/cgi-bin/cancer/datasets.cgi> (accessed on 15 December 2018).
56. Faris, H.; Mafarja, M.M.; Heidari, A.A.; Aljarah, I.; Zoubi, A.M.A.; Mirjalili, S.; Fujita, H. An efficient binary salp swarm algorithm with crossover scheme for feature selection problems. *Knowl.-Based Syst.* **2018**, *154*, 43–67. [[CrossRef](#)]
57. Wang, T.H.; Li, W. Kernel learning and optimization with Hilbert–Schmidt independence criterion. *Int. J. Mach. Learn. Cybern.* **2018**, *9*, 1707–1717. [[CrossRef](#)]
58. Yager, R.R. Entropy measures under similarity relations. *Int. J. Gen. Syst.* **1992**, *20*, 341–358. [[CrossRef](#)]
59. Sun, L.; Liu, R.N.; Xu, J.C.; Zhang, S.G.; Tian, Y. An affinity propagation clustering method using hybrid kernel function with LLE. *IEEE Access* **2018**, *6*, 68892–68909. [[CrossRef](#)]

